VOLUME IX MAY, 1918 NUMBER 5

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

AN ABSOLUTE POINT SCALE FOR THE GROUP MEASUREMENTS OF INTELLIGENCE. PART I.*

ARTHUR S. OTIS

Surgeon General's Office, Washington, D. C.

CONTENTS

Part I.
I Introduction: Purpose.
II The Tests.
Requirements of a scale for mass testing.
Description of the tests.
III The Preliminary Investigation.
IV Acquisition of the Data.
Administration of the tests.
Scoring.
V The Reliability of the Scale.
Probable errors of the test scores.
Reliability coefficients.
VI Graduation of the Scale.
Theoretical considerations.
Equating the scores.
Weighting and combining the scores.
Age norms.
Completing the Absolute Point Scale.
Coefficients of Brightness.
Part II. (See June number)
VII Overlapping of Ability between Grades.
VIII Refinement of the Scale.
The order of difficulty of the test elements.
The diagnostic value of the single test elements.
IX Inter-test Correlations.
X Further Considerations regarding Reliability.
The Reliability Coefficient of the Point Scale.
The Probable Error of the Scale.
XI Comparisons with School Mark and Amount of Schooling.

Appendix I. Sample Extracts of Tests.

*The writer is indebted to Dr. Lewis M. Terman of Stanford University for many helpful suggestions during the making of this study.

Appendix II. Showing the Point Scores of Each Pupil in Each Test.
Appendix III. Some Mathematical Reasoning with Regard to Criteria of Tests of Intelligence.
Appendix IV. Inter-Test Correlations (Raw and Corrected).
References.

I. Introduction

Purpose—The purposes of this study are:

(1) To construct a scale for the measurement of general mental ability, such scale being:

(a) suitable primarily for administration to the pupils in grades 4, 5, 6, 7, and 8 of the elementary school,

(b) capable of being administered to *groups* of at least 50,

(c) so constructed that the scoring is both rapid and, as far as possible, free from the error of the personal factor,

(d) built upon the general plan for an Absolute Point Scale outlined by the writer (see Ref. 10). Such construction would involve the "validation" and "graduation" of the tests, the determination of the probable error of a determined measure of general mental ability, etc.

(2) To investigate the correlation of the mental abilities tested by the scale.

It is not deemed feasible, in the space to which this article is limited, to attempt to discuss the nature of intelligence, such as it is presumed to measure by the present scale, nor the various definitions of intelligence which have been given or may be implied by the various 'intelligence scales' in present use. These subjects will be touched upon in various connections in the discussion. The writer will therefore proceed immediately to describe the manner in which the present scale was constructed.

II. The Tests

Requirements of a Scale for Mass Testing.—The chief object of testing in groups, of course, is economy of time. One of the most essential means of accomplishing this purpose is that the responses required be very simple. This makes for speed both in the administration and in the scoring of the tests. It has been the aim, therefore, so to arrange the tests in the present scale that there would be only one correct answer to each item and that this might be indicated merely by making a letter or figure or drawing a line. Where convenient, provision was made for the responses to be placed in a single column in which case the papers may be scored with dispatch

by the use of scoring forms. When every answer is either right or wrong, a large amount of time is saved that might be necessary otherwise to determine the value of partially correct answers. Moreover, under these conditions the tests must be scored in the same way by all investigators, this assuring comparability.

The ideas for the tests have been derived from various sources, chiefly, perhaps, from the Stanford Revision of the Binet Scale. (see Ref. 14). The general character of most of them is no doubt familiar.* In substance the remaining tests were designed especially for this study.

Description of the Tests.—The scale was compiled in duplicate. There were, in other words, two complete tests of each kind. The two tests of each kind were made as nearly alike as possible without using the same material. In each scale the tests were constructed as follows:

The Spelling Test† consisted of fifty pairs of words in two columns. The words of each pair consisted of the correct and incorrect spelling of a single word. In some cases the first spelling was the correct one and in some cases the second was the co rect one. The pupil was required to indicate by the letters, F, S, or N, placed in a parenthesis opposite the words, as shown in the sample in Appendix I, whether the first or the second was the correct spelling, nor neither spelling was correct.

The Arithmetic Test consisted of 16 problems in which the computation was made as easy as possible and the emphasis thus placed upon reasoning.

The Synonym and Antonym Test consisted of 50 pairs of words as shown in the sample. The pupils were required to indicate by the letters, S and O, whether the words of a pair meant the same or the opposite.

The Proverb Test consisted of 20 proverbs in two sets of ten proverbs, each set followed by twelve statements, one of which "explained" each of the ten proverbs, there being two extra statements in each set not explaining the proverbs. The pupils were required to place in the parenthesis before each proverb the number of the statement which explained it.

*Thanks are due to Mrs. Mary D. Chamberlain for the Proverb Test used in this study. The words of the Spelling Test, as explained later, were taken from Ayres' list. (See Ref. 2.)

†Although the Spelling Test was found, in both the preliminary and in the present investigations, to afford quite as good a measure of intelligence, from one point of view, as the other tests (see intercorrelations), it has since been considered best to drop this test from the scale.

The Disarranged Sentence Test consisted of 26 sentences with the words disarranged, as shown in the sample. The pupils were required to rearrange the words mentally to make sense and indicate whether the sentences so constructed were true or false by underlining the words true or false at the end of the line.

The Relation Test consisted of 24 items, each in the form of a proportion in which one of the four terms was to be supplied, indicating by number from five alternative answers given on the same line.

The Geometric Test consisted of 22 items, using as a basic principle that described by Abelson. (See Ref. 1.) Referring to the figures constructed by overlapping one or more circles, triangles, and rectangles, the pupils were required to place figures 1, 2, etc., in certain designated spaces as suggested in the sample.

The Following Directions Test consisted of 14 problems requiring the pupils to place certain numbers in certain figures on the Woodworth and Wells Cancellation Sheet. This test presumably differed from the preceding one in that its difficulty consisted more in the comprehension of involved language while in the Geometric Test, the difficulty lay chiefly in tracing out the space relations.

The Narrative Completion Test was of the type used by Whipple, Ebbinghaus, Terman, and others. It consisted of a short story of which certain words were omitted leaving blanks in which the pupils were required to write the words which in their judgment best fitted into the story.

III. The Preliminary Investigation

To aid in choosing the tests for the Point Scale and in determining the most suitable forms in which to give them, a preliminary investigation was conducted in which 29 pupils in grades 4 to 8 of a small school were tested. Fifteen tests were used. Eight of these were in the same or nearly the same form as those shown in the Appendix. Two others, the Arithmetic and Spelling Tests, have since been entirely made over. The Synonyms Test was given orally first, then about three weeks later repeated in the form shown. The other five tests were: a test in word meaning recognition, of the type suggested by the writer (See Ref. 8) and called the Reading Test; a test in the reproduction, in writing, of sentences dictated, called the Memory for Sentences Test; the Trabue Completion Test (see Ref. 15); the Kansas Silent Reading Test (see Ref. 4); and the Starch Grammar Test (see Ref. 12). The list is shown in Table I. The tests marked with the asterisk were given in duplicate, the double scores being used in the correlations.

The correlation of each test with the composite of all scores except those of the Oral Synonyms and Grammar Tests are shown in the first column of the table. The correlations with mental age as determined by the Stanford Revision of the Binet Scale are shown in the second column. Considering the small number of individuals as well as the unreliability of scores, the coefficients may be regarded as of suggestive value only.

TABLE I.

Some Results of the Preliminary Investigation

Test	Correlation with Composite	Correlation with Mental Age
Relation	.94	.97
Proverbs*	.94	.94
Following Directions	.86	.95
Geometric*	.89	.92
Trabue Completion Test*	.88	.88
Reading*	.92	.82
Kansas Silent Reading Test	.90	.88
Synonyms (Oral)	.79	.87
Synonyms (Written)	.83	.85
Disarranged Sentences*	.86	.81
Narrative Completion	.86	.80
Arithmetic	.84	.80
Spelling*	.79	.84
Memory for Sentences	.77	.82
Memory for Digits*	.72	.42
Starch Grammar Test	.30	.49
Correlation of Mental Age with Composite Score:	.94	
Same, "corrected for attenuation" (estimate):	.99	

IV. Acquisition of the Data

Administration of the Tests.—The tests were given to 121 children of a large grammar school—43 in the fourth grade, 40 in the sixth grade, and 38 in the eighth grade. Each test was taken by all of the pupils of one grade at a time in their regular room, the teacher being present. The writer personally conducted all the tests, giving all the directions and explanations. The tests were given in approximately the order indicated in the foregoing section. In most instances two test series were given to each grade each day, one in the morning and one in the early afternoon. The giving of the first and second tests of the same kind were separated by three or more days in most instances but particularly in the cases of the Arithmetic, Geometric, and Following Directions Tests, in

*Double scores used.

which the second test is in the nature of a recast of the first. Time was allowed for all to finish in nearly all cases, except, of course, the Disarranged Sentence Test, which is a speed test. Occasionally when one or two pupils lagged far behind the others, their papers were taken up before they finished. In such cases it was usually noted that the pupils had permitted themselves to be distracted from their work. In the Disarranged Sentence Test sufficient time was allowed for only one pupil to finish. The order, "Stop," was then given and the time noted. For purposes of comparison, all scores were afterward increased to a five minute basis.

The pupils of each grade were adjured at the beginning of the testing not to give or receive aid during the taking of any tests. A wholesome attitude appeared to be taken by all during the testing. In such instances of apparent collusion as were noted, the pupils were quietly cautioned. These instances were few. On the whole the pupils were orderly and attentive and signified their interest in the testing.

Scoring.—In the case of each test except the Synonyms, Spelling, and Disarranged Sentences, one count was given for each correct answer and no count for incorrect or omitted answers. In the case of Synonyms, however, since there are but two alternative answers, S or O, theoretically, of the answers given concerning the pairs of words not known by any pupil, but guessed at, one half will be right by chance. Therefore, if say 35 of the 50 were known and correctly marked, and 10 of the remaining 15 guessed at, leaving 5 blank; of the 10 guessed at, 5 might be marked rightly by chance. This would make 40 correct, 5 incorrect, and 5 blank. It seemed, therefore, that as many counts should be deducted from the total correctly marked (40) as were incorrect (5) thus giving a score of $40-5=35$, the number assumed to be known. A person guessing at all of them and getting half right by chance would then attain a score of $25-25=0$. This method was adopted in scoring the Synonyms and Disarranged Sentence Tests. The case of the Disarranged Sentences is complicated by the fact that sentences wrongly marked on account of haste are penalized additionally by the loss of time in scanning them. The suitability of the method, therefore, should perhaps be investigated.

Since there are three possible answers, F. S, or N, in the Spelling Test, theoretically $\frac{1}{3}$ of the number of those guessed at would be marked rightly by chance. This would mean that, to follow the

above method, the score should be obtained by deducting from the number rightly marked, ½ the number of those wrongly marked. However, inasmuch as there were only fourteen individuals who did not attempt all the words, and to avoid possible negative scores, the scores were obtained by giving one count for each right answer, no count for each wrong answer, and ⅓ count for each blank, on the assumption that if guessed at, ⅓ of these words would have been rightly marked, This brought all the scores to the same basis and necessitated counting only right answers in all but 14 cases. Identical rank orders of the individuals are obtained from the scores by the two methods. The scores that would have been obtained by using the first method can easily be derived from those used merely by multiplying by 1½ and subtracting 25.

In order to obtain a suggestion as to the value of the above methods of scoring, the sum of the differences between the first and second scores of the 14 pupils above mentioned was found first when the scores were obtained as above and second when obtained by merely counting the number of correct answers, taking no account, therefore, of the element of chance. The sum of the differences in the first case was 34 and in the second 45, although the scores were less in the second case. This suggests that the method employed was the more reliable.

While there is, of course, a 'one-in-five chance' of an element of the Relation Test being marked rightly by guess, it was not deemed necessary to take account of it. In an auxilliary investigation regarding the scoring of the Digit Test, the papers were scored (1) according to the number of digits in the last number correctly reproduced, (2) according to the number of digits in the next to the last number correctly reproduced, and (3) according to the last group of numbers of the same size of which two or more were correctly reproduced. No one of these three methods appeared to be appreciably superior to the others. The reliability coefficient of scores by method (1) was .53, by the method employed in this study, .74. It was discovered after giving the test that some of the pupils were able to reproduce numbers of nine or more digits within three trials. If such had been included in the test, the reliability coefficient by method (1) would no doubt be higher. It is believed, therefore, that with a sufficiently exhaustive test, the loss in reliability of method (1) would be more than made up by the great saving in time of scoring.

Briefly, the plans of scoring were as shown in Table II.

TABLE II.

Summary of Plans of Scoring

TEST	SCORE
Spelling	1 count for each correct answer and ⅓ count for each blank. (nearest whole number)
Arithmetic	1 count for each correct answer.
Synonyms	1 count for each correct answer and 1 count deducted for each incorrect answer (blanks not counted).
Memory for Digits	1 count for each number entirely correct.
Proverbs	1 count for each correct answer.
Disarranged Sentences	1 count for each correct underlining with 1 count deducted for each incorrect underlining.
Relation	1 count for each correct answer
Geometric	1 count for each figure 1 correctly placed, provided no other figure 1 appeared in the same design, and similarly, 1 count for each figure 2 correctly placed.
Following Directions	1 count for each direction correctly followed.
Narrative Completion	1 count for each blank satisfactorily filled.

The scores for each individual in each test will not be given as obtained by the above plan but instead they will be given in an altered form explained below. The scores are given in Appendix II.

V. THE RELIABILITY OF THE SCALE

The reliability of a test may be expressed in two ways, either (1) by giving the probable error of a score in the units of the score, the probable error being the value of that error which is exceeded in amount by half the errors, or(2) in terms of the coefficient of correlation between two tests of the same kind. The probable error of a score as a measure of the reliability of a scale is comparable with other values of the probable error found in connection with the testing of other groups of individuals but it is not comparable with the probable error of the scores of other tests unless the units in the two scales measure the same increments of ability. This would not happen often and only accidentally. The reliability coefficient, as has been explained more fully elsewhere (see Ref. 11), derived from measures of one group is not comparable with a reliability coefficient for the same test derived from measures of another group unless the heterogeneity of the two groups is the same or nearly the same. In many instances, of course, this is not the case. The reliability coefficient of one test, however, is comparable with that

of another test when both are derived from measurements of the *same group*. The reliability coefficients are necessary under these conditions to show the relative reliabilities of two tests. They compensate the measures of reliability for inequalities of scale units.

We have therefore found both the probable errors and the reliability coefficients of each of the ten tests. The probable errors of scores in the several tests were found according to the method described at length in Ref. 11. This method is expressed by the formula,

$$\text{P. } \epsilon. = \frac{\text{Med. Dif.}}{\sqrt{2}}$$

in which Med. Dif. is the median difference between scores by the same individuals in the two tests of the same kind. (A test of the first scale is called Test I; the corresponding test of the second scale Test II.) Before making the subtractions, however, it is necessary to have the scores of both tests in terms of either one or the other of the two tests, since these are quite often somewhat different; due to slight differences in difficulty, to practice effect, etc. For the purpose of evaluating the scores of one test in terms of the other, plots were made in which the scores in Test I were represented as abscissae (horizontally) and those of Test II as ordinates. The manner in which the scores in the two tests corresponded was then found by drawing in each plot a line of relation. This is such a line that the abscissa and ordinate of any point on it represent corresponding scores in the two tests. By inspection of the plots, it was deemed valid to draw a straight line of relation in all cases except that of the Narrative Completion Test, in which it was apparent that the true line of relation was markedly curved. In that case the curve of relation was drawn by the method we have called the method of correspondence by rank (see Ref. 7). In all cases except that of the Narrative Completion Test, the line of relation was obtained by finding the means, M_x and M_y, of the values of x and y and the average deviation, A. $D._x$ and A. $D._y$., of the distributions of values of x and y, and then drawing a line through the point (M_x, M_y) having a slope such that the tangent of the angle formed with the X axis $= \frac{\text{A. D.}_y}{\text{A. D.}_x}$

To find the score in terms of Test II which corresponds to any given score in terms of Test I, it is necessary merely to find the point on the relation line corresponding to the score in Test I and to note the

score in Test II at the left which corresponds to this point. The differences between the scores, in terms of Test II, are measured by the distances of the points of the plot above or below the line; in terms of Test I, by the distances to the right or left. The values of *Med. Dif.* were obtained by the method. (See Ref. 7):

$$\text{Med. Dif.} = .8453 \times \text{Avg. Dif.}$$

That is,

$$\text{P. E.} = \frac{.8453\ (\text{Avg. Dif.})}{1.414}$$

The values of the probable errors of each of the several tests were obtained first in terms of Test II and the corresponding values in terms of Test I were derived by dividing by the tangent of the angle of the line of relation. The values of the probable error in both terms are given in Table III.

TABLE III.

Reliability of the Tests

	Probable Errors		Reliability Coefficients	
	Scale I.	Scale II.	Single Tests	Double Tests
1. Spelling	1.49	1.45	.942	.970
2. Arithmetic	.74	.80	.871	.931
3. Synonyms and Antonyms	1.96	1.86	.753	.85)
4. Memory for Digits	1.04	1.21	.746	.855
5. Proverbs	1.17	1.02	.761	.864
6. Disarranged Sentences	1.28	1.76	.737	.849
7. Relation	1.50	1.86	.729	.843
8. Geometric	1.22	1.15	.805	.892
9. Following Directions	.75	.97	.82)	.901
10. Narrative Completion	5.43	3.80	.840	.913
			Total	8.877

The formula used for finding the reliability coefficients was:

$$r = 1 - \tfrac{1}{2}\left(\frac{\text{A. D.}_{(\text{dif's})}}{\text{A. D.}_{(\text{scores})}}\right)^2$$

which is a variation of the difference formula:

$$r = 1 - \tfrac{1}{2}\left(\frac{\sigma_{(y-x_y)}}{\sigma_y}\right)$$

This latter formula is the equivalent of the Pearson product-moment formula. (See Ref. 11.) In these formulae, $\text{A. D.}_{(\text{dif's})}$ and $\sigma_{(y-x_y)}$ are measures of the variability of the distribution of

differences between the scores of each of the 121 pupils in Test I and Test II, when the scores in Test I are evaluated in terms of Test II; and in which A. D.$_{(scores)}$ and σ_y are respectively corresponding measures of the variability of the distribution of *scores* in Test II.

The reliability coefficients thus found for each test are shown in the third column of Table III.

We were quite surprised to find the Spelling Test to be so much in the lead in this rating. However, the Spelling, Narrative Completion, and Synonym Tests had 50 elements while the other tests had only 25 or less. The Arithmetic, Following Directions, and Geometric Tests no doubt have an advantage over the others in that Test II was only slightly different from Test I.

The aim in duplicating the tests, as has been stated, was to make the second test in each case as nearly like the first as possible without actually copying it. This was done in order that the score in the second test would be as near as possible to a second score in the same test. It is possible that a second score in the same test would have been preferable for finding the reliability if it had been convenient to separate the two givings of the tests by a sufficient interval. Even this, however, would introduce new sources of error. Since the differences in difficulty between the two tests of a kind are not the same for all the pupils, the differences between the scores in the two tests tend to be greater than would be the case it the same test could be given twice, even without memory of the first testing, in which case the difference in the scores would be due merely to differences in disposition at the times of taking the first and second tests. For this reason, the values of the probable errors and reliability coefficients, considering only errors due to varying disposition, are really less than those given here.

Further consideration regarding reliability will be given later. These depend upon the values of inter-test correlations.

VI. Graduation of the Scale

Theoretical Considerations.—There are two aspects to the graduation of the scale. One deals with the proper combining of the scores of the several tests and the other with the finding of age norms, percentage norms, etc. The scores of each individual in the ten tests must first be combined into a single score, say a "point-score," and then those point-scores may be determined which are

normal for each of the given ages of childhood, or those which given percentages of adults may be expected to attain, etc.

In order that the scores of an individual in the several tests may be properly averaged, it is necessary to take account of the differences in value of the units of the scales of the several tests. If an increment of one problem in an Arithmetic score is in reality equal to an increment of four words in the score of the Synonym Test, to average the scores in the two tests just as they stand would be to give the Synonym Test four times as much weight as the Arithmetic Test. If, therefore, it is desired to give equal weight to each test, the score of an individual in each test must be transmuted into other terms, say "points," such that *equal increments of ability* in each test receive *equal increments of points*. It is convenient, also, while assigning point values to the scores in the several tests, to arrange that correspondihg *amounts* of ability in the several tests shall receive corresponding *numbers* of points. The first of these conditions is essential and the second convenient for the purpose of averaging scores properly; both conditions are essential for the purpose of comparing scores in the several tests with one another. If it seemed reasonable to assume that for the individuals of a given group, the ability possessed by the upper 25% in any test was as much above that possessed by the upper 50% as that ability was above the ability possessed by the upper 75%, and if the ability in any one test was considered equal to that in any other test which was possessed by the same percentage of individuals, then the first of the above mentioned conditions would be complied with by representing the difference between upper 25% and upper 50% ability in each of the several tests by some number of points (say 10), and the difference between upper 50% and upper 75% ability in each of the several tests by that *same* number of points (10). And the second condition would be complied with by representing 50% ability in all of the ten tests by the same number of points (say 50) in which case, of course, 25% ability would be represented in each case by 60 points and 75% ability by 40 points.

We have been speaking of the equality of increments of ability, but such equality is a very indefinite thing. Equal increments of ability must be such as are measured by the same number of units of some kind. We have not been willing to grant that the steps of any test scale necessarily measured equal increments of ability.

Nor would we admit that any year's growth in ability is equal to every other year's growth. The growth of ability is supposed to retard eventually with age. In what units then will we say ability may be measured so that equal numbers of units measure equal increments of ability? In a previous article (Ref. 10) we have suggested that absolute units of ability be so defined that the distribution of abilities of all adults *will be normal* (in the technical sense). This would mean that those percentages of adults which were considered as possessing abilities which marked successive steps on an absolute scale of ability were the same percentages as those of the normal probability surface which corresponded to successive units of the base. Until such time, however, as a very large number of unselected adults have been tested, such a criterion of equality of units of ability will be unavailable. In lieu of such a criterion, an alternative method was used.

The Procedure Used for Determining Equality of Increments of Ability.—Although we have felt that the units in one part of a single test scale were very apt to be of greater value than those in some other part, it is quite probable that if the upper units of some test scale must be considered as measuring greater increments of ability than the lower units, the opposite probably might be considered true of some other test scale, so that taking the test scales all together, the median value of the units in one part may be considered as equal to the median value of the units in any other part. Proceeding upon that hypothesis, the most probable true form of the distribution of abilities of the 121 pupils was determined by obtaining a composite of the separate distributions for the ten tests as follows:

1. The score attained in each test by the 30th individual in rank (beginning with the lowest) was assigned a preliminary point-value of 40 points and the score attained by the 90th individual in rank was assigned a preliminary point-value of 60 points.*

2. Tentative point values corresponding to all the other scores were then determined in such a manner that the units in all parts of the test scale were represented by equal increments of points. This was accomplished graphically in each case by drawing a straight line.

3. From the smooth curves of distribution of test scores were

*These scores were not the *actual* scores of those individuals but the scores corresponding to them on smooth curves through the distributions of consecutive scores.

then determined the scores attained by the 3rd, 9th, 15th, 60th, 105, 111th, and 117th individuals in rank order. These points in the distribution curves were believed to best reveal any skewness of the distribution.

4. The preliminary point values corresponding to the scores attained by the 3rd individual in each test distribution were then ascertained. These were then plotted in order of magnitude and a median value determined by means of a smooth curve through the plotted points. This median point value was 24.4. The other median point values were as follows:

Individual in order:	3,	9,	15,	(30)	60,	(90)	105,	111,	117
Point value	24.4	29.7	33.3	(40)	50.1	(60)	66.7	70.1	75

It should be stated that these values indicate that the distribution of abilities of the 121 pupils approximately normal.

5. Since the median of the preliminary point values obtained by the 3rd individual in rank in the several test distribtuions was 24.4, this value may be assumed to be the most probable true value, in terms of our established absolute units, of the ability in any test which the 3rd individual in rank order attained. The score in each test attained by the 3rd individual in rank order (by the curve) was then given, therefore, the corrected point value 24.4. Similarly the score in each test attained by the 9th individual was then given the corrected point-value 29.7, etc.

6. In order to determine the corrected point value to be similarly assigned to all the other scores in each test, a graph was made for each test in which the preliminary point values corresponding to the scores attained by the 3rd, 9th, etc., individuals were plotted as ordinates and the new point values, 24.4, 29.7, etc., plotted as abscissae. A smooth curve was then drawn through the series of plotted points. This curve was then taken as showing the relation between the preliminary and corrected point values corresponding to each score in the test From this curve for each test were taken the corrected point values corresponding to each score. These are shown in Table IV. They no doubt represent the nearest approach that can be made to a true absolute point scale.

Considerations with Regard to Weighting and Combining the Scores.—After finding the corrected point values corresponding to each test score, the scores of each pupil in each test were transmuted into terms of points and the total score found for each. These are given in Appendix II.

This method of combining the scores resulted in equal weight being given to each test. No doubt some of the tests are more significant than others in the measurement of general ability, however we conceive it. Unreliability of a test, of course, lowers its significance. Other aspects of significance depend upon the conception of general ability. If a test is considered as measuring general ability only to the extent to which the factors entering into the ability tested are common to other abilities, both as to number of factors and as to number of abilities to which they are common, then the degree to which a test may be considered as measuring

TABLE IV

Showing the Number of Points Corresponding to Each Score in Each Test

Score	Spelling Points	Arith. metic Points	Synon- yms Points	Digits Points	Proverbs Points	Disar. Sentences Points	Relation Points	Geomet. Points	Fol. Direc. Points	Narra. Comple- tion Points
0		21		20	32	23	25	21	23	26
1		24		20	34	28	26	22	27	26
2		27		21	35	33	27	24	31	27
3		30		22	36	38	28	25	35	27
4		33		23	37	43	29	27	39	28
5		36		24	39	47	31	28	43	28
6		39		26	40	50	32	29	48	29
7		42		28	41	53	33	31	52	29
8		45		31	43	56	35	32	56	30
9	20	48		35	44	58	37	34	60	30
10	21	51		38	45	61	40	36	64	31
11	22	54		42	47	63	42	38	68	31
12	23	56		45	48	66	45	41	72	32
13	24	59		48	49	68	47	43	76	32
14	25	62		51	50	71	50	46	80	33
15	26	65		55	52	73	52	49		33
16	26	68		58	53		55	52		34
17	27			61	55		57	56		34
18	28		20	64	56		60	59		35
19	29		22	68	57		62	63		35
20	30		23	71	59		65	66		36
21	31		25				67	70		36
22	32		26				70	74		37
23	32		28				72			38
24	33		29				75			38
25	34		31				77			39
26	35		32							39
27	36		34							40
28	37		35							41
29	38		37							41
30	38		38							42
31	39		40							43
32	40		41							44
33	41		43							45
34	41		44							46
35	42		46							47
36	43		47							48
37	44		49							49
38	45		50							50
39	46		51							51
40	47		53							52
41	48		54							53
42	49		56							55
43	50		57							56
44	51		59							57
45	52		60							58
46	53		62							59
47	55		63							61
48	56		65							62
49	58		66							63
50	60		68							64

general ability is expressed by the amount of "correlational spread" of the test, to use McCall's expression, by which is meant the sum of the intercorrelations of the test with other tests comprising a fairly representative collection, each presumed to involve factors common to the others. The last qualification is necessary since, if the group of tests is too restricted in kind, certain 'specific' abilities may be common to too large a proportion of the tests and thus vitiate the criterion of general ability.*

On the other hand if a test is considered as contributing to the measure of general ability if it measures an ability that may be considered valuable in aiding the individual to adjust himself to the new problems and conditions of life, whether such ability has few or many factors in common with others; then it is not proper to use only the criterion of correlational spread. Two possible alternatives suggest themselves. If there were available for the individuals tested a satisfactory criterion of their powers of adaptation to the new conditions and problems of life, in the nature of a measure of economic or scholastic success, then it would be necessary merely to weight the tests according to the regression equation method, so as to obtain the best correlation of the composite score with the criterion. In lieu of such a criterion, the tests might be weighted according to a combination of the weights assigned by a number of judges. In this study, for instance, the results of all the tests except that of Memory for Digits correlated uniformly highly with each other. The Digit Test, which showed a reliability not the least among the ten tests, stood quite apart from the other tests in showing low correlations with all of them. According to the criterion of correlational spread, this test would be weighted very much lower than any of the others. According to either of the criteria pertaining to the second conception of general ability, however, the Digit Test might perhaps deserve a weight more nearly the amount of the others.

---

*Some mathematical reasoning bearing on this point is given in Appendix III, 1 and 2.

McCall used this criterion in his study (Ref. 5). Another criterion which he also used was the correlation of each test with "Composite," a measure obtained by combining the scores of all the tests (with some exceptions) after weighting each according to *a priori* considerations as to the value of the tests. Although the correlations of the several tests with Composite appear to have been determined by McCall by separate calculations, it would have been possible to obtain the values of these correlations with Composite more simply from the values of the inter-test correlations. The necessary procedure is given in Appendix III, 3.

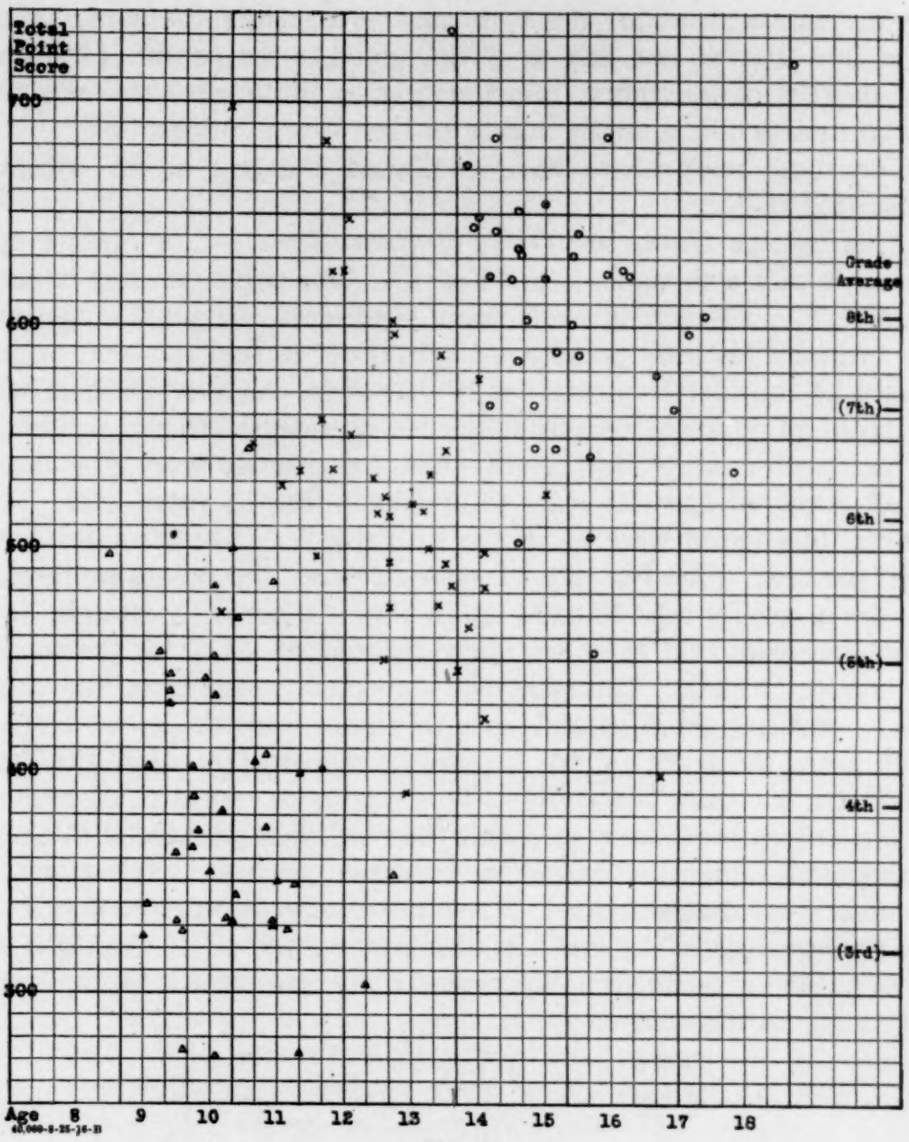


FIG. 1

Showing the Relation between Total Point Score and Age

In this study we are inclined more to the second conception of general ability mentioned. It was not feasible in this study, however, to use either of the criteria appropriate to this conception. To weight the tests according to reliability alone, it would be necessary to weight each inversely in proportion to the square of the probable error (the probable errors being in comparable terms). (See Merriman, Ref. 6, p. 95.) Such procedure, however, prac-

tically implies that all the tests aim to measure the same thing. But since they do not, any weighting given to compensate for different degrees of reliability, necessarily also emphasizes the effect of certain particular abilities and is to that extent undesirable.

For these reasons we have combined the test scores without weighting them.

Finding Age Norms in Terms of Point Scores.—For finding age norms, a plot was made. (See Fig. 1.) One point pertains to each pupil. The abscissa of each point represents the pupil's age and the ordinate his total point score. In order to find the score which would be considered normal for 10-year-olds, the average score was found of all pupils of ages from 9 years, no months, to and including 11 years, no months; for 11-year-olds, the average score was found of all pupils of ages 10 years to and including 12 years, etc. The norms thus found were as shown in Table V. These values were then plotted. (See Fig. 2.) To our surprise, the points representing the norms for ages 10 to 14 lay in almost a perfectly straight line, which suggests that they are fairly reliable, at least, for the school population tested. This was not expected considering the gaps left by omitting the fifth and seventh grades from the group tested. The norms for years 15, 16, 17 may be seen to fall below the line, the latter two quite markedly. This was to be expected, of course, since the pupils of these ages were selected, being retarded in their schooling. While the true norms for these ages are doubtless above the average values obtained, it was not deemed proper to continue the straight line. The line was therefore curved off to the right as shown. We must regard the norms for the ages above 15, as being only roughly approximate.

TABLE V.

Showing Age Norms in Point Scores

Age:		8	9	10	11	12	13	14	15	16	17	18	19
Point:	Observed:			404	446	487	527	566	583	550	584		
Score:													
Norms:	Smoothed:	324	364	405	445	486	526	566	600	624	638	647	650

Completing the Absolute Point Scale.—We have previously (see Ref. 10) given the name, Coefficient of Brightness, to the quotient that would be obtained by dividing the measure of the absolute amount of mental ability of any individual by the measure of the absolute amount of mental ability which was normal for the age of that individual. This means, of course, that the measures of

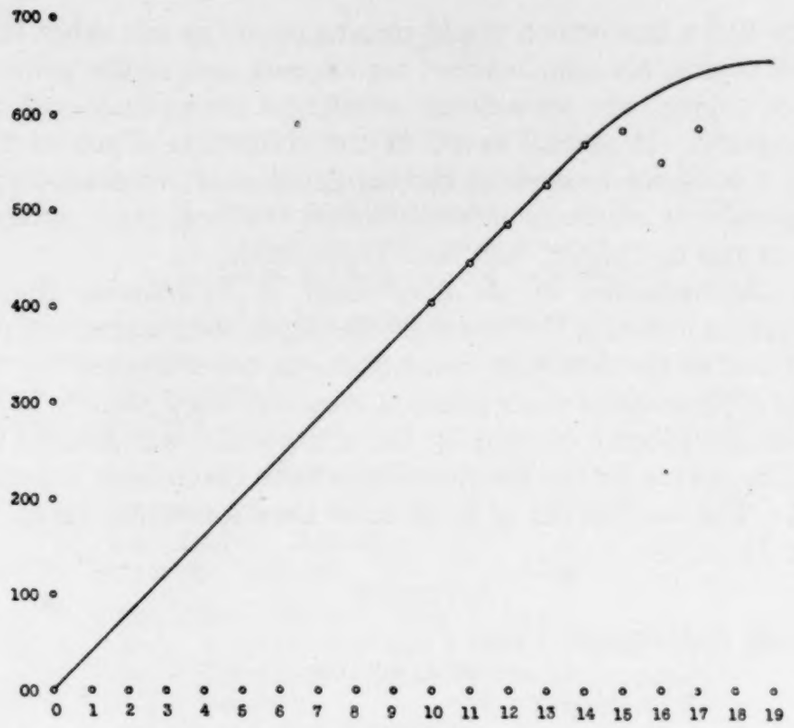


FIG. 2
Showing a Smooth Curve through the Age Norms of Total Point Scores

mental ability must be in such terms that not only will equal increments of ability be measured by equal increments of the scale, but twice as many units on the scale will represent twice as much ability, etc. In other words, zero of the scale must represent just absence of ability. Before it was possible for us to find the coefficients of brightness of the pupils tested in this case, therefore, it was required to note what correction was necessary in the scale of points in order that the number of points representing the ability of age 0 would be 0. The ages for which we may presume to have obtained fairly reliable norms are only those from 10 to 14. Inasmuch, however as the increments of points between the norms for these ages are almost exactly the same, it was regarded as proper to assume for present purposes that if continued, the line through the norms would be straight the rest of the way to age zero. It was then necessary to note what number of points thus corresponded to age zero, this number to be considered the true absolute zero of the final point scale. To our further surprise, it was discovered that by calling the yearly increment of points (below 14) approx-

imately 40.5 a line, which would pass as nearly as any other through the five norms, actually reached zero age at zero of the point scale. This, of course, was an entirely accidental coincidence and not at all necessary. It merely saved us the obligation of subtracting or adding a constant to each of the corrected point values assigned to the several test scores in order to obtain the final point values constituting the completed Absolute Point Scale.

The Determination of the Coefficients of Brightness.—Since the point values in which the scores of the pupils were expressed proved to be those of the Absolute Point Scale, in order to find the coefficients of brightness of each pupil, it was necessary merely to divide the total point score of each by the score which was normal for his age. The norms for the fractional ages were taken from the curve in Fig. 2. The coefficients of brightness thus found are given in Appendix II.

APPENDIX I.

Sample Extracts of Tests:

TEST 1: SPELLING

1. forenoon	fournoon	(F)
2. intrest	interest	(S)
3. neighber	neighbor	( )
4. concider	consider	( )
5. entertain	entertane	( )
etc.	etc.	etc.

TEST 2: ARITHMETIC

1. If a boy has 10 cents and then earned 5 cents, how much did he have then? ..... ( ) cents

7. How many years will it take a glacier to move 1000 feet at the rate of 100 feet a year? ..... ( ) years

15. A ship has provision to last her crew of 50 men 6 months. How ong would it last 30 men? ..... ( ) months

TEST 3: SYNONYMS AND ANTONYMS

1. large	big	(S)
2. decrease	increase	(O)
3. empty	vacant	( )
4. knowledge	ignorance	( )
50. conservative	radical	( )

TEST 4: MEMORY FOR DIGITS

1. 4739	( ) ( ) ( ) ( )
2. 2854	( ) ( ) ( ) ( )
3. 7261	( ) ( ) ( ) ( )
4. 31759	( ) ( ) ( ) ( ) ( )
5. 42385	( ) ( ) ( ) ( ) ( )
6. 98157	( ) ( ) ( ) ( ) ( )

TEST 5: PROVERBS

Pr verbs

(3) Make hay while the sun shines.
( ) In a calm sea every man is a pilot.

Statements to explain the proverbs.

1. Deeds show the man.
2. Leadership is easy when all goes well.
3. Make the best of your opportunities.

TEST 6: DISARRANGED SENTENCES

1. name a John is boy's.....(*true* false)
2. sun morning the the in sets.....(true *false*)
3. trees birds nests the in build.....(true false)

TEST 7: RELATIONS

hand : arm : : foot : ( )
2. hat : head : : thimble : ( )
23. education : ignorance : : ( ) : poverty

1. 1 leg, 2 toe, 3 finger, 4 wrist, 5 elbow.
2. 1 finger, 2 needle, 3 thread, 4 hand, 5 sewing.
23. 1 laziness, 2 school, 3 wealth, 4 charity, 5 teacher.

TEST 8: GEOMETRIC TEST

(*Designs were presented composed of two or more geometrical figures—circles, rectangles, and triangles—overlapped.*)

1. Place a figure 1 so that it will be both in the rectangle and in the circle.

7. Place a figure 1 so that it will be in both circles, in the triangle, and in only one rectangle.

TEST 9: FOLLOWING DIRECTIONS

(*A page of Woodworth and Well's Cancellation Test was supplied each pupil.*)

2. In line 1 [of the forms] place a figure 1 in the first star and a figure 2 in the second circle.

7. In line 5, place a figure 7 in the form which follows the same kind of form as that which follows it.

TEST 10: NARRATIVE COMPLETION

Once upon a ..... there was a y..... ..... who was very p..... He went from place to ..... trying to find .....

APPENDIX II.

Showing the Point-Scores of each individual in each test.

Eighth Grade

Pupil	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16
Spell.	77	55	68	56	63	64	59	62	65	57	54	70	72	53	63
Arith.	76	54	67	49	51	76	56	58	73	56	72	60	64	44	62
Synon.	76	58	64	50	57	67	72	59	57	60	56	75	69	47	64
Digit.	80	76	41	45	43	62	53	50	41	50	62	32	45	30	36
Proverb.	81	54	65	35	46	62	65	54	68	60	52	68	71	35	62
D. Sen.	68	72	63	63	68	75	63	72	60	59	71	53	72	59	63
Rel'n.	61	53	71	44	53	49	69	69	67	69	67	75	73	41	63
Geom.	52	55	58	55	58	58	66	69	74	66	64	58	61	45	61
Fol. D.	67	58	65	48	56	51	67	61	58	72	58	75	75	43	56
Compl.	80	62	64	58	68	58	65	68	68	72	48	68	83	56	58
Sums	718	597	626	503	563	622	635	622	631	621	604	634	685	453	586
C. B.	109	93	100	86	89	108	108	99	103	103	102	108	110	74	96

Pupil	17	18	19	20	22	23	24	25	26	27	28	29	30	32	33
Spell.	64	70	67	65	51	64	56	64	54	72	65	72	61	54	56
Arith.	66	58	49	64	58	64	70	56	66	64	70	56	64	73	51
Synon.	58	67	65	64	56	61	55	54	65	68	68	59	54	66	65
Digit.	48	62	50	60	45	43	30	57	76	57	41	66	30	64	57
Provb.	52	71	60	65	52	52	52	54	60	81	65	58	56	68	50
D. Sen.	67	63	60	70	68	64	53	73	65	63	68	66	71	75	68
Rel'n.	67	61	55	75	46	61	55	53	46	61	71	42	60	69	65
Geom.	55	69	52	64	64	40	49	66	66	61	64	61	64	55	72
Fol. D.	58	67	38	58	53	53	58	53	63	73	61	61	48	65	73
Compl.	54	63	50	56	43	63	64	55	44	72	69	60	57	66	66
Sums	589	651	546	641	536	565	542	585	605	672	642	601	565	655	623
C. B.	98	111	91	104	84	98	88	99	95	120	111	99	95	109	100

Fourth Grade

Pupil	34	35	36	37	38	39	40	41	51	52	53	55	56	58	59
Spell.	64	64	64	54	68	51	70	64	27	32	39	31	53	46	29
Arith.	58	60	67	77	70	67	70	79	44	27	39	34	19	46	39
Synon.	50	53	74	48	72	63	46	73	19	35	32	32	25	43	29
Digit.	55	45	75	75	53	36	64	78	41	36	68	50	45	50	36
Provb.	46	50	68	56	56	52	71	75	40	31	35	24	31	31	31
D. Sen.	49	59	68	70	68	65	65	63	48	38	40	25	48	49	42
Rel'n.	58	46	61	56	60	75	75	82	26	32	46	29	30	38	38
Geom.	29	52	72	66	61	58	66	72	29	36	41	43	25	45	36
Fol. D.	43	61	72	53	70	51	58	75	29	29	46	41	29	41	32
Compl.	54	56	62	66	66	60	64	71	27	36	44	35	35	52	38
Sums	506	546	683	621	644	578	649	732	330	332	430	344	340	441	350
C. B.	83	92	119	107	115	92	115	133	75	87	113	82	93	110	79

Pupil	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
Spell.	36	40	40	43	30	41	46	53	37	29	31	35	39	38	37
Arith.	27	23	39	32	36	32	39	49	36	23	44	36	44	46	32
Synon.	43	30	29	37	32	47	62	48	50	31	29	27	40	54	40
Digit.	57	27	34	41	55	57	48	45	57	50	50	53	43	69	30
Provb.	35	31	31	24	35	48	68	56	40	31	31	24	45	50	40
D. Sen.	38	27	32	37	41	35	46	41	45	27	40	40	40	34	34
Rel'n.	44	28	25	31	12	55	44	44	33	25	49	38	56	51	34
Geom.	45	22	38	22	28	49	55	55	23	30	45	41	58	45	38
Fol. D.	32	21	38	29	35	46	48	48	38	35	41	29	43	51	43
Compl.	31	22	23	30	24	59	41	61	45	22	39	26	43	47	38
Sums	388	271	329	326	328	469	497	500	404	303	399	349	451	485	366
C. B.	99	67	73	90	85	112	145	120	94	61	87	77	111	110	93

Pupil	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89
Spell.	45	32	42	40	36	50	33	46	35	34	44	29	44	51	15
Arith.	41	27	27	32	36	46	36	32	44	39	34	36	49	49	34
Synon.	50	34	27	28	43	51	28	30	49	33	28	39	36	54	46
Digit.	34	24	64	64	45	57	39	39	64	55	22	22	43	45	50
Provb.	48	38	24	35	31	43	40	43	52	31	38	40	38	35	48
D. Sen.	45	25	28	30	43	54	32	45	45	38	23	32	40	45	44
Rel'n.	42	36	32	34	41	53	31	46	55	39	32	33	38	42	29
Geom.	61	52	25	38	33	64	38	41	45	26	43	45	52	41	37
Fol. D.	43	29	29	43	46	61	41	41	51	35	29	46	51	32	35
Compl.	34	37	33	37	48	66	35	39	42	33	39	33	43	42	36
Sums	443	334	331	381	402	545	353	402	482	363	332	355	434	436	374
C. B.	117	80	75	93	102	127	68	110	119	95	80	88	107	115	94

Sixth Grade

Pupil	90	91	92	94	95	96	101	102	103	104	105	107	108	109	110
Spell.	42	8	34	28	33	31	35	50	52	54	52	49	58	40	64
Arith.	41	36	34	32	32	51	62	62	54	44	49	44	49	49	51
Synon.	48	22	43	22	37	34	74	56	45	45	54	50	59	46	59
Digit.	60	30	50	29	39	73	50	60	55	24	69	29	50	62	48
Provb.	40	24	35	35	35	43	65	62	43	40	62	38	45	43	52
D. Sen.	45	41	51	19	45	32	47	40	48	40	41	56	53	45	47
Rel'n.	44	29	41	31	35	38	55	60	58	42	60	53	46	38	58
Geom.	43	28	32	24	49	41	45	66	55	31	45	45	52	38	49
Fol. D.	41	32	43	29	38	32	58	75	56	25	56	43	61	46	51
Compl.	49	23	38	26	32	33	44	71	59	45	45	39	72	58	50
Sums	453	273	401	275	375	408	535	602	525	390	533	446	545	465	529
C. B.	122	59	85	71	85	93	117	117	87	75	99	80	99	83	118

Pupil	111	112	114	115	118	119	120	121	122	123	124	125	127	128	129
Spell.	55	44	61	47	51	70	36	41	56	47	51	59	56	68	56
Arith.	51	44	54	39	51	56	56	46	41	44	51	58	79	49	49
Synon.	49	40	62	51	41	75	43	50	55	54	53	45	72	51	52
Digit.	62	45	69	50	27	75	57	55	41	32	53	45	57	68	60
Provb.	58	56	75	56	38	71	56	56	50	45	48	58	65	46	52
D. Sen.	40	45	49	56	51	55	33	41	61	45	48	26	56	47	48
Rel'n.	44	53	58	48	23	69	48	55	46	56	60	71	73	58	55
Geom.	45	38	81	49	41	77	55	43	47	52	47	69	81	61	49
Fol. D.	48	48	70	38	32	78	56	61	46	51	58	53	72	51	65
Compl.	45	37	69	39	44	72	54	52	56	45	48	48	61	52	58
Sums	497	450	648	473	399	698	494	500	499	471	517	532	672	551	544
C. B.	106	88	132	92	64	168	96	93	87	115	102	106	144	112	127

Pupil	130	131	132	133	134	135	136	138	139	140	141	142	143	144	145	146
Spell.	64	55	53	45	64	46	52	47	44	47	51	49	40	52	61	59
Arith.	49	64	69	60	41	39	51	54	58	46	39	51	62	67	51	56
Synon.	73	62	62	55	48	43	48	43	53	55	56	55	52	48	63	64
Digit.	43	50	34	64	60	53	69	34	75	62	48	53	30	53	48	62
Provb.	81	65	62	58	54	40	40	52	60	65	60	58	46	45	71	71
D. Sen.	54	58	52	46	49	38	49	45	41	65	44	43	46	49	52	57
Rel'n.	61	60	55	58	48	39	48	56	60	55	49	36	58	58	49	67
Geom.	72	69	52	52	49	38	58	66	69	35	40	41	49	66	52	49
Fol. D.	65	51	56	43	58	43	61	41	58	43	53	38	51	38	70	80
Compl.	63	63	63	55	53	44	44	46	60	42	53	51	48	41	70	60
Sums	625	597	558	536	524	423	520	484	578	515	493	475	482	517	587	625
C. B.	131	115	118	112	102	74	99	88	101	100	90	87	85	97	108	129

Concluded in the June Number

LOGICAL MEMORY AND SCHOOL GRADES

IRVING KING AND T. B. HOMAN

University of Iowa

It was the purpose of this study to determine what sorts of tests of immediate memory for sense material show the best if any correlations with scholastic marks. But few investigations of this problem have been reported and the results are fragmentary and conflicting. The most extended report is by A. Travis, in *The Psychological Clinic* of December, 1915, "Reproduction of Two Short Prose Passages: A Study of Two Binet Tests." Travis found that these two short prose passages from the Binet tests, when given to groups of college students, and immediate reproduction is called for, not only do not correlate in the results with one another but do not show any correlation with class marks. On the basis of these results Travis raises the question as to whether other memory tests of a similar character but of different material might be devised which *would* correlate with other abilities. The writers of this article undertook a further investigation of this question. They sought to determine whether the *length* of the prose passage used as a test is of material importance for the test and whether the *kind* of material used is an important element. The different types of material selected may be roughly described as *historical, descriptive,* and *narrative.* Each type of material was represented by tests of three different lengths, a short passage which required from one to one and one-half minutes for the experimenter to read, a medium length passage, requiring from two to three minutes for the reading and a slightly longer passage requiring from four to six minutes for the reading. Special effort was made to secure passages in which the thought was clear-cut. Copies of the passages used will be furnished any one desiring them. For the sake of brevity only the results are reported here.

The subjects available for these tests were 64 pupils in the 5th to the 8th grades of the University of Iowa Elementary School, 72 pupils in the University High School and 110 Junior and Senior students in the College of Liberal Arts.

A different set of paragraphs was used with the High School and University students from what was used with the pupils of the elementary school but all groups had the *short, medium* and *long* tests of the three types, except that the College group took only the narrative paragraphs. The elementary and high school groups therefore each took nine *immediate* memory tests and the college group three. Each group took also the same number of *deferred* memory tests, being called upon to give a second reproduction of each passage after a lapse of from 24 to 48 hours. The results of the deferred reproductions were also compared with scholastic standings.

These tests were all given by the regular teachers to the various class groups which, in most cases, consisted of about 35 pupils or students each. Definite and uniform instructions for conducting the tests were followed by all the teachers. After the first day, before a new paragraph was presented, the subjects wrote what they could recall of the preceding test paragraph.

In addition to the scholastic marks which were to be available for these groups, a series of "mental tests" marks were secured for the high school groups. These pupils, early in the first semester of the year 1916–17 were given a part of the tests used with the University Freshmen, viz: Courtis Arithmetic, Series B, three paragraphs of Simpson's Completion test; a set of twenty harder mixed relations or analogies; the three inch cube test; an opposites test (40 of the easier words in Simpson's hard lists; see the standardizations reported by Gold and King, Jour. Ed. Psych., October, 1916); and logical memory, (The Dutch Homestead). We therefore had two separate measures of ability for the high school group.

All the memory tests were graded on the basis of thought units of the original passages reproduced, full credit being given for all reproduced thought units whether in the same or in different words.

All correlations were computed on the basis of ranks in scholarship and in tests, the Pearson formula adapted to rank differences being used.

Results with Elementary Subjects

The following table gives the correlations computed between the various immediate and deferred memory tests and the scholastic ranks of the elementary school group.

TABLE I.

CORRELATION TABLE — ELEMENTARY GROUP

Showing Coefficients of Correlation Between the Several Memory Tests and the Scholastic Rank

	"r's"
Scholastic ranks and His. Short Im.	.35
" " " " " Def.	.37
" " " " Med. Im.	.36
" " " " " Def.	.55
" " " " Long Im.	.39
" " " " " Def.	.44
Average	.41
Average of immediate memory test correlations	.37
Average of deferred memory test correlations	.45
Scholastic ranks and Desc. Short Im.	.52
" " " " " Def.	.48
" " " " Med. Im.	.34
" " " " " Def.	.32
" " " " Long Im.	.29
" " " " " Def.	.38
Average	.42
Average of immediate memory correlations	.40
Average of deferred memory correlations	.43
Scholastic ranks and Nar. Short Im.	.37
" " " " " Def.	.36
" " " " Med. Im.	.12
" " " " " Def.	.47
" " " " Long Im.	.38
" " " " " Def.	.35
Average	.24
Average of immediate memory correlations	.29
Average of deferred memory correlations	.39
Average of all Short	.42
" " " Medium	.37
" " " Long	.37

It will be noted from this table that the average correlations of the second reproductions in each of the three groups of passages are slightly higher than the correlations of the immediate reproductions. The tests with the descriptive material yield slightly better correlations than do the other groups of material and the short passages yield a higher average correlation than do the medium and long.

Results with High School Subjects

The ranks of the high school group based on the combined mental tests were correlated with scholastic ranks, yielding an r = .54.

Table II gives the correlations between "mental test" ranks, scholastic ranks and memory ranks.

TABLE II.

High School Group — Correlation Table

Showing Coefficients of Correlation Between the Logical Memory Tests and the Combined Intelligence Rank

				"r's"
History	Short	Im.	Mental test rank	.42
"	"	Def.	" " "	.41
"	Med.	Im.	" " "	.29
"	"	Def.	" " "	.09
"	Long	Im.	" " "	.32
"	"	Def.	" " "	.31
			Average	.30
Average of immediate memory correlations				.34
Average of deferred memory correlations				.27
Descriptive	Short	Im.	Mental test rank	.30
"	Def.		" " "	.25
"	Med.	Im.	" " "	.5
"	"	Def.	" " "	.54
"	Long	Im.	" " "	.34
"	"	Def.	" " "	.43
			Average	.40
Average of immediate memory correlations				.39
Average of deferred memory correlations				.40
Narrative	Short	Im.	Mental test rank	.23
"	"	Def.	" " "	.33
"	Med.	Im.	" " "	.30
"	"	Def.	" " "	.36
"	Long	Im.	" " "	.41
"	"	Def.	" " "	.36
			Average	.33
Average of immediate memory correlations				.31
Average of deferred memory correlations				.35

Average for all Short ..... .32
" " " Medium ..... .35
" " " Long ..... .36

Here as in the case of the Elementary School group the results from the descriptive passages yield the highest correlations, and in two of the sets the second reproduction gives better results than the first.

Table III gives the correlations of the various tests with average scholastic ranks of the High School group.

TABLE III.

HIGH SCHOOL GROUP CORRELATION TABLE

Showing Coefficients of Correlation Between the Logical Memory Tests And the General Scholastic Rank

					"r's."
History	Short	Im.	Scholastic	ranks	.27
"	"	Def.	"	"	.20
"	Med.	Im.	"	"	.27
"	"	Def.	"	"	.21
"	Long	Im.	"	"	.19
"	"	Def.	"	"	.29
				Average	.23
Average of immediate memory correlations					.29
Average of deferred memory correlations					.32
Descriptive	Short	Im.	Scholastic	ranks	.12
"	"	Def.	"	"	.09
"	Med.	Im.	"	"	.55
"	"	Def.	"	"	.44
"	Long	Im.	"	"	.20
"	"	Def.	"	"	.45
				Average	.31
Average of immediate memory correlations					.29
Average of deferred memory correlations					.32
Narrative	Short	Im.	Scholastic	ranks	.15
"	"	Def.	"	"	Neg.
"	Med.	Im.	"	"	.22
"	"	Def.	"	"	.25
"	Long	Im.	"	"	.34
"	"	Def.	"	"	.29
				Average	.19
Average of immediate memory correlations					.22
Average of deferred memory correlations					.16
Average for all			Short		.13
"	"	"	Medium		.32
"	"	"	Long		.29

Table IV gives the correlations of the various tests taken by the college group with their average scholastic ranks.

TABLE IV.

CORRELATION OF LOGICAL MEMORY WITH SCHOLASTIC RANK.

CORRELATION TABLE

Juniors and Seniors, College of Liberal Arts

				"r's"
Narrative	Short	Im.	Scholastic rank	.06
"	"	Def.	" "	.08
"	Med.	Im.	" "	.47
"	"	Def.	" "	.43
"	Long	Im.	" "	.43
"	"	Def.	" "	.45
			Average....	.32

Average of immediate memory correlations..... .32
Average of deferred memory correlations..... .32

From these tables it seems that the longer passages are better tests than the shorter ones, the coefficients of correlation between the grades in these tests and scholastic standings being fairly high. Other comparisons of the tests with the different groups of subjects may be easily made by inspecting the tables. The most significant fact is perhaps that nearly all the correlations are fairly high and that only one is negative. On the whole the medium length passages yield the highest average correlations for all groups taken together, namely, .38. The average correlation of the short passages was .29 and of the long was .35.

The average correlation for the historical passages was .32; for for the descriptive .38 and for the narrative, .34.

In Table V is given the median percentage of ideas remembered by the different scholastic groups.

TABLE V.

Median Percentages of Ideas Reproduced (immediate recall) by the Different Groups

GROUP	MEDIAN PERCENTAGE OF IDEAS REMEMBERED
Elementary.....	25.
High School.....	32.
College, L. A.....	39.

Table VI gives the percentages of subjects in the different groups who had higher average standings in the deferred recall than in the immediate.

TABLE VI.

Percentages of Subjects in the Different Groups Who Had Higher Average Results in the Deferred Recall than in the Immediate

GROUP	PERCENTAGE OF SUBJECTS WHO GAINED IN THE DEFERRED RECALL
Elementary	19.
High School	40.
College, L. A.	45.

GENERAL CONCLUSIONS

1st. As the pupils advance in scholastic grade their power to remember and immediately reproduce prose passages of the type here used increases.

2nd. The pupils' ability to reproduce such prose passages after an interval of 24 to 48 hours increases more rapdily with scholastic grade than does their ability to reproduce them immediately.

3rd. All coefficients of correlation obtained in these various processes were positive with the exception of the one between logical memory, narrative type, short selection, deferred recall and the general ability of the high school group.

4th. The elementary group showed the highest coefficient of correlation between the general scholastic ranks and the ranks found in the historical material, medium length, deferred recall.

5th. The high school group showed the highest coefficient of correlation between the general scholastic ranks and the ranks in the descriptive material, medium length, immediate recall.

6th. The college group showed the highest coefficient of correlation between the general scholastic ranks and the ranks in the narrative material, medium length, immediate recall.

7th. The memory correlations are highest in the case of the elementary school pupils.

8th. The averages indicate that there is little difference between immediate and deferred memory, in this type of test, as a measure of scholastic ability.

9th. The averages indicate that a medium length passage as a test of logical memory is the best measure of scholastic ability except for the elementary pupils.

10th. The averages also indicate that the descriptive type of prose passages correlate best with the general scholastic standing. The narrative is second, and the historical third.

11th. The coefficients of correlation secured by comparing the results of the tests in logical memory with the general scholastic ranks, indicate that a logical memory test may be considered a fair measure of the general scholastic standing of the individual student.

12th. There seems to be no evidence of a practice improvement in any of the groups.

THE MEASUREMENT OF PROGRESS IN LANGUAGE ABILITY

RUDOLF PINTNER
Ohio State University

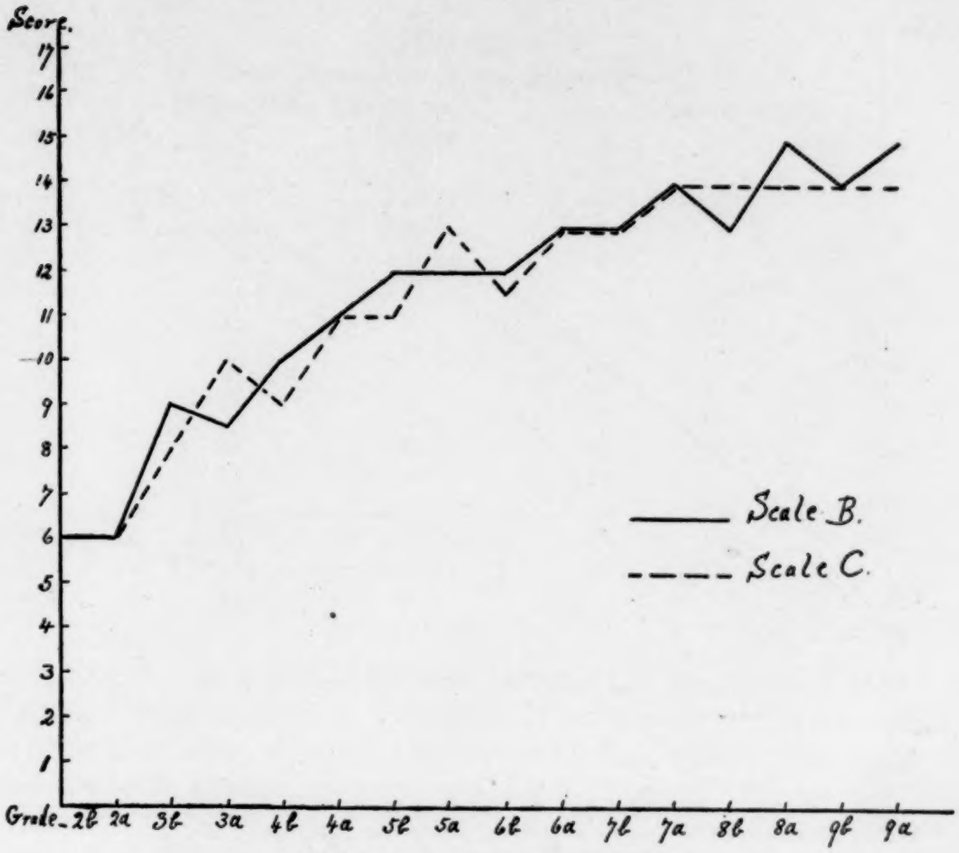
The Trabue Language Completion Tests*, Scales B and C, were used to obtain a measure of the increase in language ability of school children. Two schools were tested; the one a junior high school with an enrollment of about 750, and the other an ordinary grade school with an enrollment between 550 and 650. The junior high school would be considered a rather superior school, whereas the other would be ranked as an average city school.

Scale B was given to the children in October, 1916, and Scale C in May, 1917. No change in the method of language instruction was made by the teachers and for the most part they were not aware of the purpose of the tests, because several other tests were given at the same time. At the time of the first test they did not know that the pupils were to be re-tested at a later date.

The procedure in giving and scoring the tests followed the standard method prescribed by Trabue.

The grade medians for each school are shown in Graphs I and II. These medians are for the grades irrespective of the children in those grades, that is to say, the median for the 2a grade is not based upon the same children for Scale B as for Scale C, since the majority of the pupils in 2a have advanced to 3b in May when Scale C was given. The graphs show the language ability for each grade in the school and unless any radical change in language teaching has taken place, we would not expect to see any decided difference between the two scales. Graph I shows the results for the junior high school and Graph II the same for the grade school. According to expectation we do not see any decided difference between the curves for the two scales. The fluctuations are such as might arise owing to the differing abilities of the children making up the separate classes. The two schools are shown on separate graphs because the overlapping of the four curves on one graph would be too confusing. A comparison of the two graphs, however, will show that on the whole the grade school (Graph II) is slightly lower than the junior high school (Graph I).

*TRABUE M. R. *Completion Test Language Scales.* Teachers College Contributions to Education, No. 77, New York, 1916.



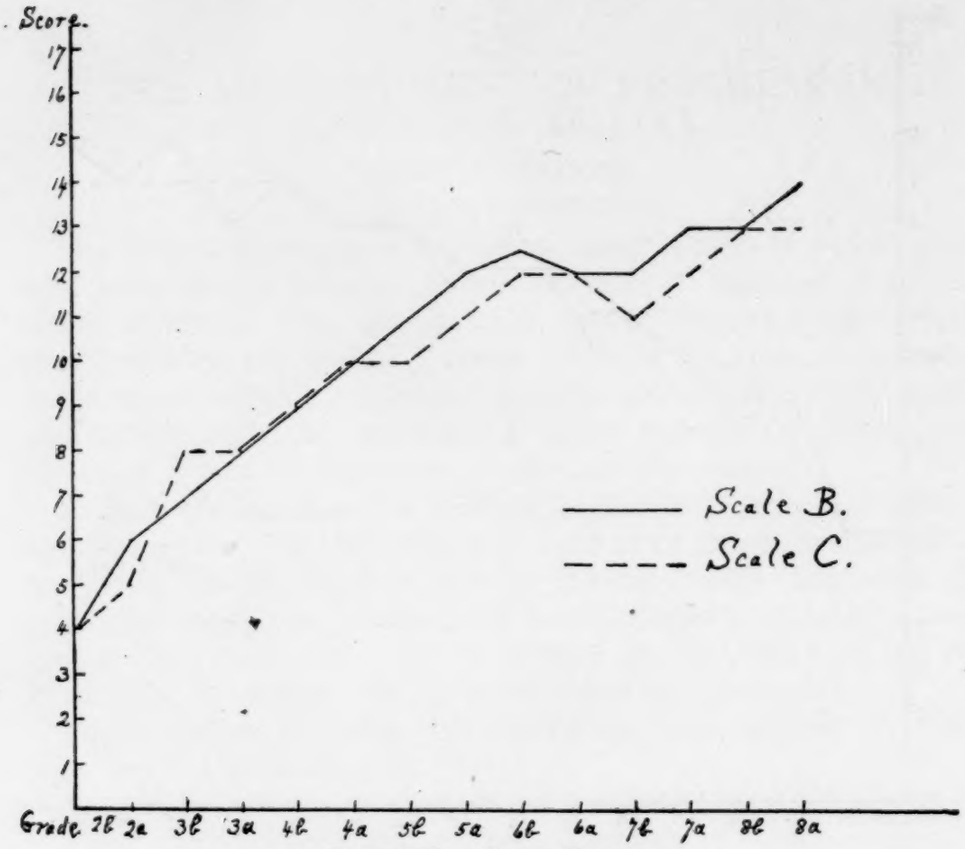
GRAPH I. JUNIOR HIGH SCHOOL

Table I shows the medians for each grade, irrespective of the

TABLE I.

Grade Medians for Scales B and C

	JUNIOR HIGH SCHOOL		GRADE SCHOOL	
	B.	C.	B.	C.
Grade	October	May	October	May
2	6	6	5	4
3	8.75	9	7.5	8
4	10.5	10	9.5	9.5
5	12	12	11	11
6	12.5	12.25	12.25	12
7	13.5	13.5	12.5	11.5
8	14	14	13.5	13
9	14.5	14		
Number of Cases.....	751	746	558	634



GRAPH II. GRADE SCHOOL

different sections of each grade. These medians give us the normal grade ability in language for the two schools. In both cases there is very little difference between Scales B and C, showing that the scales are roughly equal in difficulty. Only in two cases is there a difference of one point, namely in the second and seventh grades in the grade school, where one point less was scored on Scale C than on Scale B. This may result either from the greater difficulty of Scale C, mentioned by Trabue, or from a poorer group of children in those grades in May as compared with October. On the whole, however, the table seems to show that a school keeps a certain standard of language ability from year to year.

A first rough comparison of relatively the same group of children can be obtained from comparing each grade on the B Scale with the next higher grade on the C Scale, because the children as a whole have progressed from the one grade to the higher grade during the period covered by the tests. This comparison is shown in Table II. The table is to be interpreted as follows: In the junior

TABLE II.

Grade Medians for B and C Compared

	Junior High School			Grade School		
Grade	B.	C.	Difference	B.	C.	Difference
2b–2a	6	6	0	4	5	+1
2a–3b	6	8	+2	6	8	+2
3b–3a	9	10	+1	7	8	+1
3a–4b	8.5	9	+0.5	8	9	+1
4b–4a	10	11	+1	9	10	+1
4a–5b	11	11	0	10	10	0
5b–5a	12	13	+1	11	11	0
5a–6b	12	11.5	−0.5	11	11	0
6b–6a	12	13	+1	12	12	0
6a–7b	13	13	0	12.5	12	−0.5
7b–7a	13	14	+1	12	11	−1
7a–8b	14	14	0	12	12	0
8b–8a	13	14	+1	13	13	0
8a–9b	15	14	−1	13	13	0
9b–9a	14	14	0			
			Ave. = +0.47			Ave. = +0.32

high school the 2b grade in October became the 2a grade in May, and made a median score of 6 on Scale B in October and a score of 6 on Scale C in May, showing a difference of zero, while in the grade school the same grade increased from 4 to 5 showing a difference of plus one; the next group of children (2a to 3b) showed the same medians and the same increase in both schools. It will be noted that on the whole the lower grades show some increase in the median, whereas the upper grades show hardly any. The steady increase in the median seems to stop at the fourth grade. Above this the greater number of grades show a difference of zero. The average of all the differences shown in columns four and seven shows a gain of +0.47 for the junior high school and +0.32 for the grade school.

Eliminating now all individuals who did not take both tests, we find that we have 598 in the junior high school and 418 in the grade school. Tabulating the individual gains or losses in score from Scale B to Scale C, and averaging these gains or losses for each grade, we find the results shown in Table III. In the junior high school in the group of 36 children that moved from 2b to 2a the average gain in score was +2.31 points and the median gain +2; in the corresponding group in the grade school the gain was +0.81 and the median gain zero, and so on for the rest of the table. Again we note that there is a constant and steady gain in the lower classes up to and including grade 4, but from there on the change from Scale

TABLE III.

Individual Gains or Losses

Grade	Junior High School			Grade School		
	Average	Median	No. Cases	Average	Median	No. Cases
2b—2a	+2.31	+2	36	+0.81	0	21
2a—3b	+2.26	+2	39	+1.50	+2	34
3b—3a	+1.79	+2	38	+0.53	+1	28
3a—4b	+0.67	+1	27	+0.61	0	28
4b—4a	+2.00	+2	40	+0.64	0	33
4a—5b	−0.25	0	24	−0.10	0	40
5b—5a	+1.06	+1	34	−0.24	0	51
5a—6b	−0.79	−1	28	+0.36	0	33
6b—6a	+0.29	+0.5	28	−0.06	0	34
6a—7b	−0.03	0	33	+0.18	−1	22
7b—7a	+1.29	+1	48	−0.38	−0.5	2[illegible]
7a—8b	+0.08	0	36	−0.77	−1	35
8b—8a	+0.87	0	63	−0.36	0	33
8a—9b	−0.97	−1	33			
9b and a mixed	−0.73	−1	41			
9b—9a	+0.08	+0.5	50			
		Total	598			418

B to Scale C fluctuates around zero, sometimes being positive and sometimes negative. Again we note that the junior high school is superior to the grade school in amount of progress made.

The distribution of all the cases, irrespective of grade, according to the number of points gained or lost is shown in Table IV. The table shows that in the junior high school two individuals, or 0.3 per cent. of the total, made a gain of 9 points (*i. e.* the score on Scale C was 9 or more than the score on Scale B); and, further, third line, that 4, or 0.7 per cent. of the junior high cases, made a gain of 7 points, and one of the grade school children made this gain, that is, 5 of all the cases under consideration, or 0.5 per cent., made a gain of 7 points. The rest of the table is to be interpreted in the same way. A comparison of the two schools is well brought out on Graph III. The junior high shows a greater number of cases making some kind of a gain and the grade school shows a greater number at zero and below. The combined curve for the two schools would, of course, run between the two curves on the graph and would show a very normal distribution.

From Table IV we can compute the average and median gain or loss for all the children for the two schools separately and combined:

TABLE IV

Distribution According to Points Gained or Lost

Gain or Loss	Junior High		Grade School		Both Schools	
	No.	Percent.	No.	Percent.	No.	Percent.
+9	2	0.3			2	0.2
+8			2	0.5	2	0.2
+7	4	0.7	1	0.2	5	0.5
+6	8	1.3	9	2.1	17	1.6
+5	21	3.5	10	2.4	31	3.0
+4	50	8.4	16	3.8	66	6.5
+3	63	10.5	36	8.6	99	9.8
+2	84	14.1	45	10.8	129	12.7
+1	86	14.4	55	13.2	141	13.9
0	91	15.2	78	18.7	169	16.7
−1	68	11.4	56	13.4	124	12.2
−2	54	9.0	56	13.4	110	10.9
−3	37	6.2	25	6.0	62	6.1
−4	16	2.7	21	5.0	37	3.6
−5	6	1.0	1	0.2	7	0.6
−6	1	0.2	7	1.7	8	0.7
−7	4	0.7			4	0.4
−8	3	0.5			3	0.3
Totals	598	100.1	418	100.0	1016	99.9

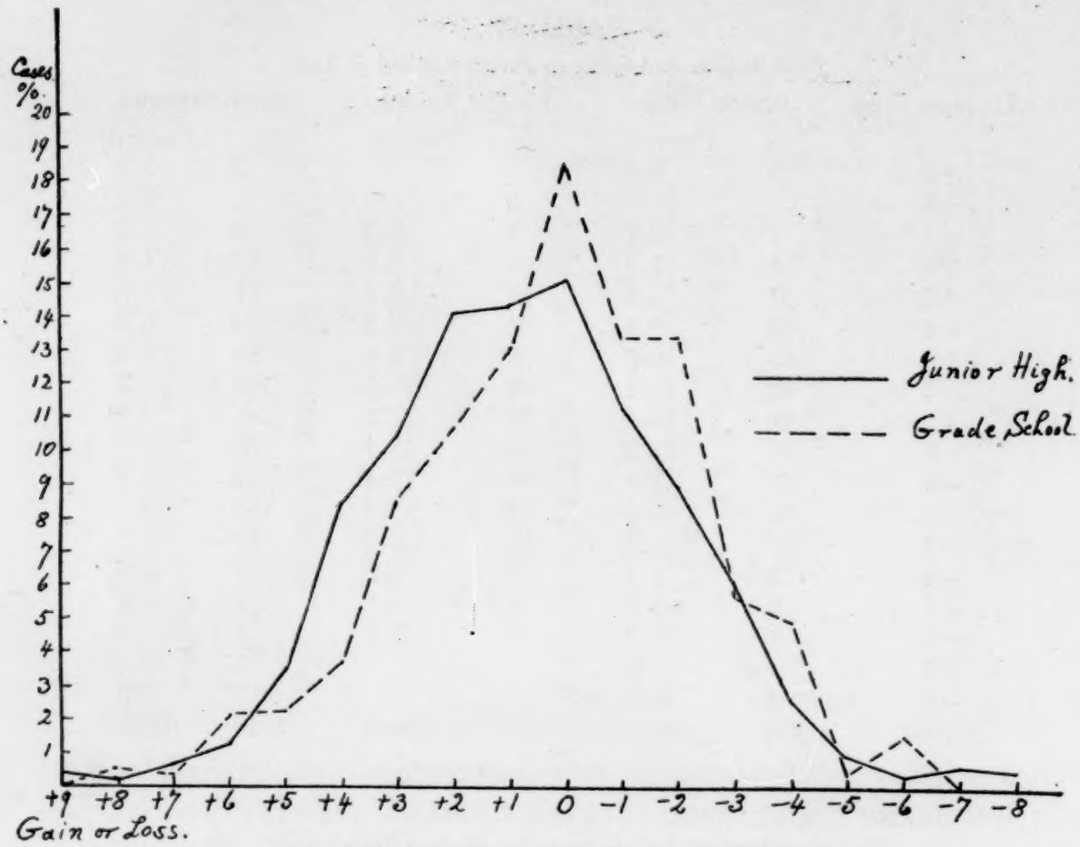
	Average	Median
Junior High School	+0.67	+1
Grade School	+0.17	0
Both Schools	+0.46	0

The individual gain is so slight that it does not show itself in the median of all the cases together. It is on the average half of one point.

We have noticed, however, all along that the marked increase was in the lower grades and that the loss in ability as measured by the scale occured in the upper grades. We have, therefore, computed the gain in score for the cases from the second through the fourth grade (October standing). This gives us the following results:

	Average Gain
Junior High School	+1.87
Grade School	+0.92
Both Schools	+1 48

With these younger children we have a gain of about one and one-half points. In the better class school it is about two points and in the other school about one. It is obvious, therefore, that the



GRAPH III. INDIVIDUAL GAINS AND LOSSES

scale is too easy for the higher grades and, therefore, does not allow of a measurement of their progress in language ability. Furthermore, the progress in language ability is very probably not so rapid as the child grows older. Since the sentences in Scales B and C are more or less so constructed as to show a difference of two points from grade to grade, we may say that in these lower grades this rate of improvement is practically taking place. A gain of 1 5 has been made from October to May leaving the additional increase of 0.5 to be made during the rest of the year, when, perhaps, language progress may not be so rapid owing to the smaller amount of formal instruction by the school during the summer months.

Finally we give below the grade medians for Scale C for all the pupils tested for purposes of comparison with other work. The grade medians for Scale B are not published here, since the writer has published elsewhere* grade medians for this scale for a much larger group of cases. The grade medians for Scale C are:

*PINTNER, R. *The Mental Survey*. D. Appleton and Company, 1918.

Grade.....	II	III	IV	V	VI	VII	VIII	IX	
Median.....	4	8	10	11	12	13	13	14	
No. of Cases..	130	180	161	204	148	170	242	145	1380

SUMMARY

1. Two schools were tested at the beginning and end of the school year by means of the Trabue Scales B and C.

2. Differences in the language ability and in the language progress of the two schools were marked and constant.

3. In both schools children above the fifth grade showed no language improvement as measured by the scales. It would seem that Scales B and C are not suitable for the measurement of language progress of older pupils during a relatively short interval of time. Harder scales may bring out better what progress may exist.

4. The language progress of the total group was about half a point.

5. The language progress of the younger children in grades four and below was 1.5 points, about what is to be expected in measuring the development for about seven months of the school year.

TESTS OF APPLICANTS FOR ADMISSION TO UNIVERSITY OF MINNESOTA MEDICAL SCHOOL

M. E. HAGGERTY

University of Minnesota

Candidates for entrance to the University of Minnesota Medical School in September, 1916, were given two tests with a view to determining mental capacity and general mental fitness to pursue the work of the school. The tests were given not so much for the purpose of accepting or excluding the immediate applicants as to provide a method and standards for the proper evaluation of future candidates. Since similar tests are being repeated with students applying for admission in September, 1917, the present paper should be regarded as a report of an investigation in progress.

No applicants were examined who did not meet the academic requirements for entrance to the Medical School. These requirements are substantially as follows: A high school course including two years of Latin, two years of prescribed and elective studies in the College of Science, Literature and the Arts in the University of Minnesota or its equivalent. During these two years of work the following studies are prescribed: rhetoric, 6 credits; physics, 8 credits; general chemistry and qualitative analysis or these subjects combined with organic chemistry, 12 credits; zoology, 6 credits; and a reading knowledge of German.

THE TESTS

Two types of tests were given. One was an omnibus test fashioned after the one reported by Scott.* This test was arranged in triplicate since the students were tested on different days and it was desired, so far as possible, to prevent communication from one group to another.

As a second test, sets 9, 9⅓, and 10⅓ of the Thorndike† Improved Scale for Reading Ability, Scale Alpha 2, were used.

METHODS OF TESTING

The candidates, 93 in all, were tested either individually or in groups ranging up to 25. Both the omnibus test and reading test

*SCOTT, W. D., *The Scientific Selection of Salesmen.* Advertising and Selling Magazine, 1916.

†THORNDIKE, E. L., *An Improved Scale for Measuring Ability in Reading. Teachers' College Record,* November, 1915, and January, 1916.

were given at one sitting. The former was given first and the paper was handed in before the reading test was given out. Accurate time was kept on both tests.

Scoring

The tests were scored both for rate and quality. In the omnibus tests each error was penalized by adding ten seconds to the raw time, thus giving a "weighted time." The three omnibus tests were rated as of the same difficulty. Subsequent studies with the same tests on approximately 2000 high school students show that there was no substantial error in so doing. In the reading tests the subjects were credited with the values achieved on the Thorndike scale as well as being scored for time. The Thorndike values are called "quality of reading" in the discussion to follow.

Results

At the time the tests were given there were no data available to show what high or low scores in either of the tests signify in terms of probable success in the work of a medical curriculum. Common sense indicates that in a course as crowded as that of the medical school the quality and rate achieved in difficult reading would be an important indication of probable success in that course. Certain experimental evidence indicates that the rate of ordinarily difficult associations is some indication of general intelligence. That the two tests measure in part the same functions is indicated by the correlation between them of .34 (weighted time in association and quality of reading. (See Table I).

This same correlation figure indicates that in part the two tests measure different things. It was thought, therefore, that a high rank in the two tests would be a double indication of superior ability and would clearly indicate the acceptability of the applicant. Low rank in both tests would argue for rejection on the basis of mental incompetence. Exception to this general rule was made in the case of two students, one Russian and one Hindu, whose language difficulties were inordinate. As a matter of fact no students were rejected wholly on the basis of the tests, as there was always "corroborative evidence" of low ability from the academic records and general information tests.

Of the 93 persons taking the tests, 69 were accepted for entrance and actually entered the medical school in September, 1917. Concerning these we have now available the records for the work of the first semester in the medical school. These records together with the pre-medical record in the academic college and the two tests are the basis of the following table of correlations.

TABLE I.

	S. L. A. Marks	Medical School Marks	Medical School Marks, 15 hrs. cre.	Medical School Marks, 14 hrs. cre.	Quality of Reading	Time of Reading	Omnsbus Test Raw Time	Omnibus Test Weighted Time	Weighted Time and Quality of Reading	Weighted Time and Time of Reading	S. L. A. and Medical School Marks	Omnibus Test, Errors
S. L. A. Marks.....		49	59	51	46	—14	39	35	55	36	85	36
Medical School Marks	49				62	24	51	60	65	57	85	29
Medical School Marks, 15 hrs. cre.....	59				77	24	68	53	81	61	88	17
Medical School Marks, 14 hrs. cre.....	51				56	4	46	47	69	43	85	45
Quality of Reading...	46	62	77	56		21	41	34	82	41	55	36
Time of Reading.....	—14	24	24	4	21		19	10	17	76	4	—18
Omnibus Test, Raw Time.....	39	51	68	46	41	19						39
Omnibus Test Weighted Time.....	35	60	53	47	34	10						67
Weighted Time and Quality of Reading.	55	65	81	69	82	17						51
Weighted Time and Time of Reading...	36	57	61	43	41	76						33
S. L. A. and Medical School Marks.....	85	85	88	85	55	4						25
Omnibus Test Errors.	36	29	17	45	36	—18	39	67	51	33	25	

For a proper understanding of this table some explanation of terms is necessary. "S. L. A. marks" means the quality of work done by the student in his college course prior to his entrance to the medical school. In the University of Minnesota the grades which a student receives carry what are called "honor points." Grades of A, B, and C merit three, two, and one honor points respectively. A grade of D or below carries no honor points. A student must have an average of one honor point per credit hour to receive a degree from the University. In computing the "S. L. A. marks" in this study the total number of honor points made by a student was divided by his number of credits. These marks therefore indicate the average quality of work done by the student in his pre-medical course. In no case was the amount of such academic work less than two full years. From this it ranged up to four years, credited with the bachelor's degree. Sixty-eight cases were available for correlation.

"Medical marks" are the marks received by the student during the first semester in the medical school. Only students carrying full work, 14 or 15 credit hours, were considered. The difference in number of credits was determined by the fact that some students took organic chemistry, a four-hour course, and others took physiological chemistry, a three hour course. There were a considerable number of irregular students; some had secured medical school credits during their pre-medical courses; others did not carry full work in the medical school during the semester in question All such students were excluded in figuring correlations. The number available for correlations were 24 with fifteen credits and 26 with fourteen credits or 50 in all.

In figuring the correlations it was desirable to credit quality as well as quantity of work. Since a very large number of students were reported with marks of D which carried no honor points at all it was clearly unfair to use the same method of figuring quality of work as was used on the "S. L. A. marks." Accordingly the following arbitrary values were assigned to the several letters:

A=10 B=8 C=5 D=2 F=0

The "medical school marks" were computed by finding the number of such points to which a student was entitled and dividing the number by the number of credits. Thus: 3 credits A=30 points, 5 credits B=40 points, 6 credits D=12 points, a total of 82 points. This total divided by 14, the number of credits, equals 58.5, the actual mark of quality to which the student is entitled. By this method the 15 credit and 14 credit students may be considered together without serious statistical error.

"Quality of reading" means score on Thorndike scale. "Time of reading" means total time required for reading test. "Omnibus Test: Raw time" equals time required for the association test. "Omnibus Test: Weighted time" is the same with 10 seconds added for each error. Available for correlation in all of these measures were 68 cases.

It was originally thought that the weighted time in the omnibus test and the quality of reading would be the best prognostic measures, and the applicants were rated in them first. As it turns out the correlation of these combined measures with the marks subsequently made by these students is .65 for the group of 50, .81 for the fifteen-credit cases and .69 for the 14 credit cases. Either of these tests alone gives a slightly lower but still substantial correlation. In no case is it less than .56.

A comparison of the correlations of S. L. A. marks and medical school marks shows that the tests were a much better indication of ability to do the work of the medical school than were the academic records of these applicants. There is not a single case in which the test correlations are so low as the academic mark correlations. In fact either test alone is a better prognostic measure than were the S. L. A. marks. The "quality of reading" correlates with medical marks to the extent of .62, the omnibus test to .60 and the two combined to .65, while the S. L. A. marks show but .49.

The prognostic value of the tests may be seen also from the per cent of median, tertile and quartile retention which turns out to be 71, 56 and 46 respectively as shown in Figures 1, 2, and 3. Thus in Figure 1, 23 persons are shown to stand in the upper half of the group in both of the tests and in the upper half of the group in medical school marks; 16 persons stood in the lower half of the group in both the tests and medical school marks. Nine persons were in the first half in the tests but in the latter half in school marks, while 2 were in the lower half in the tests but in the upper half in school marks. From Figure 3 it is apparent that no person who fell to the second or third quartile in the tests achieved a place in the first quartile in school marks. Only two in the second quartile of the tests came into the school mark first quartile. On the other hand no person who achieved a first quartile in the tests fell to the fourth quartile in school marks. But two of this test group went to the third quartile while 6 are found in the second and 10 in the first quartile. Only 9 of the 50 individuals are displaced in the marks more than one quartile from their test rankings. Such displacement as does exist is largely due to the very great number of D's given in the anatomy courses.

As in the case of the correlations, the median, tertile and quartile retentions show the tests, either singly or combined, to have been better prognostic instruments than the academic records of the students in the Arts college.

FIGURE 1.

Median Retention

RANKS IN MEDICAL SCHOOL MARKS

Standing in tests. Weighted time plus quality of reading.	Halves 1	Halves 2
1	23	9
2	2	16

Median retention 71%

These results are significant. They mean that an entrance examination board can determine by three hours work the fitness of one hundred applicants for the work in a medical school more exactly than they can derive such information from the laborious examination of academic records, often hard to obtain and equivocal in meaning.*

A curious fact is the somewhat equivocal correlations shown by the "time of reading" scores with academic marks. With the S. L. A. marks it is -14 and with medical school marks it is positive .24. From this it would seem that in a test as difficult as the reading scale the time consumed is no indication of ability. This was not true of the association test where the "raw time" scores gave positive correlations of .39 and .51 with S. L. A. and medical school marks, .68 and .46 with the 15 and 14 credit student's marks and .41 with the quality of reading. This probably means that in the simpler associations the rate is more important than in the more difficult tests and more important than the quality of work as measured by the errors in this test.

It is probably fair to interpret all these results as follows: As a means of prognosis for success in the medical school course, it is

*The giving of the two tets occupied the time of two examiners for not to exceed one and one-half hours. Since the applicants may be tested in groups, a hundred may be examined at a sitting. The scoring of the tests may be done very rapidly by persons familiar with the work.

FIGURE 2.
Tertile Retention
RANKS IN MEDICAL SCHOOL MARKS
Tertiles

Standing in Tests. Weighted time plus quality of reading	1	2	3
1	13	6	6
2	2	7	5
3	0	5	8

Tertile Retention 56%

FIGURE 3.
Quartile Retention
RANKS IN MEDICAL SCHOOL MARKS
Quartiles

Standing in tests. Weighted time plus time of reading	1	2	3	4
1	10	6	2	0
2	2	5	1	6
3	0	1	3	4
4	0	1	4	5

Quartile Retention 46%

important to know the rate and quality of fairly simple associative processes (weighted time in omnibus test) and also to know how difficult an associative problem a student can master more or less regardless of the time required to do it (quality of reading).

The desirable amount of correlation between mental tests and academic achievement is difficult to fix. Success in school work is apparently dependent upon many factors, such as good observational powers, memory, imagination and intelligence or reasoning capacity. Presumably such tests as the ones here used are directed to the measurement of such intellectual functions. Beyond these, however, are other mental functions such as industry, endurance, interest, persistence of motive, freedom from prejudice, physical fitness, general tonicity of muscular system, and a wide range of instinctive and emotional states and activities illy defined but fairly described as "attitudes," "sets," etc. It these tests measure merely the intellectual functions, then a certain portion of the student's mark is due to mental abilities not evaluated by the tests. In such cases the correlation should fall below 1, or a perfect figure. The amount it falls short would be determined by the degree to which such non-intellectual functions determine the school mark. In such cases, a correlation figure of .60, .70 or .80 might be as high as could be expected and should be regarded as the maximum to be expected from tests of this type. In case .80 were determined as such maximum, a correlation of .60 with a test for intellectual functions would be explained as due to the inadequacy of the test for the measurement of such functions. In such case it would be necessary to improve the test or to supplement it with other tests for intellectual capacities. With the fixed maximum of .80 a test showing .80 would be regarded as serving all the legitimate ends of the intellectual test. Its proper supplement would be tests for the instinctive, emotional and other factors to which a certain portion of the student's mark is due.

It is extremely doubtful, however, whether such vivisection of mental capacities is the way to their proper evaluation. Association in an individual is not the abstract thing we talk about in our text books but is shot through with the very interests and emotions we so carefully discriminate it from in our discussions. When a person attacks the "hard directions" test it is not merely the abstract processes of association, memory, discrimination, and attention which are involved. It is all of these modified by the desire to suc-

ceed, to do the task quickly, to obviate errors, and suffused with the affective and emotional states prompted by the surprise, absurdity and complexity of the test. The time score recorded at the end of the test is, therefore, a measure of the complex mental state involving all of these functions. Similarly, when an individual works diligently for 30 minutes on the Thorndike reading scale it is not fair to assume that the result of his work is merely an intellectual product. It is a mental product measuring his intellectual capacities, to be sure, but also indirectly testing his interest, his endurance, persistence and other emotional and instinctive attitudes. The implication of those considerations is that a correlation of .70 between a test and school marks means that the test measures more than the intellectual factors operating to produce the marks but measures them all partially. It means further the possibility of a test which will correlate much higher with actual performance than would be the case if the test were of purely intellectual functions. Conceivably the correlation could approach a perfect relationship. This means that none of the measures available from these tests is a satisfactory one. Valuable as they are it is probable that they may be improved so as to yield even higher correlations than the ones here reported. Two changes, on the basis of experience, seem worth while. The reading test should be presented with more degrees of difficulty so as to distribute the individuals more widely. The association test should be made of more uniform difficulty throughout. This means that some substitution should be made for the first three parts which are apparently so easy for students of the advancement of those here considered that the mechanical performance of writing the answers consumes more time than the mental processes of thinking the correct response.

In the testing of the students applying for entrance to the Medical School in September 1917 these changes have been made. The value of these changes is yet to be determined.

TESTING PRACTICE MATERIAL IN THE FUNDAMENTALS

CYRUS D. MEAD
College for Teachers, University of Cincinnati
and
CHARLES W. JOHNSON
Williams Avenue School, Norwood

One of the contributions which the science of measurements and tests makes to the field of education is the tools it offers for evaluating different methods of instruction (or learning) and different materials which enter into the teaching or learning process. Not many years ago, the only means of determining merit in material or method was by judgment or opinion, based, perhaps, upon so-called experience or formal examination. Such a judgment might vary as the convictions or prejudices of the teacher or administrator varied. The construction and application of standardized tests puts a somewhat impersonal instrument in the hands of the teachers so that the results obtained by the use of such tests are susceptible of less variable interpretation than the more familiar examination. The challenge put to any claimant in educational matters nowadays is, upon what data or results do you base your statements? The more impersonal such data are, the more weight such statements carry.

The purpose of this experiment was to obtain results upon which an opinion could be based as to the comparative degree pupils were made proficient by the use of two kinds of practice material, not only proficient in the fundamentals of arithmetic, but in the so-called arithmetical reasoning phases. The latter problem will be taken up in the second part of this study. Such an experiment as regards proficiency in the fundamentals had previously been made with nearly one thousand fifth grade pupils in the Cincinnati schools working in twenty-four classes of fourteen schools. In a few months' time, using not more than fifteen minutes daily (which included distributing and collecting of material, and recording scores), most of these pupils were advanced beyond sixth grade standards regardless of the kind of material used. The experiment complete, with tabular results, has been issued as a *School Efficiency Monograph* by the World Book Company. This large group experiment probably allowed varying factors to enter into the results, although the

conditions were controlled as nearly as possible. The direct purpose of this test then was to carry out the same experiment with smaller groups under more direct supervision and control. This problem was assigned, accordingly, to three members of a graduate seminar in Education,* each man being the principal of a large elementary school. To them chiefly is due the credit for a very painstaking and thorough piece of work.

Six classes were selected, four fifth grades and two sixth grades. They were paired so that three used one kind of practice exercises and three another kind. The material used was the Courtis Standard Practice Tests (Exercises), and the Thompson Minimum Essentials. Each principal supervised two classes, one class having one kind of material, the other, the second kind. The experiment extended over a period of about ninety school days,beginning about the first of December, 1916, and closing near the middle of April, 1917. In order to measure progress made by two groups, an identical test must be given to each group preliminary to and directly after the different practice materials are used. By comparing the median gains (or losses), or the percentile growths, one should be able to form a fairly accurate estimate of the changes due to each of the kinds of practice material used between the preliminary and final tests. The initial ability of the six classes in the fundamentals was obtained by giving the Courtis Standard Research Series B test. (See Table IV and Charts I and II.)

On the first Monday in December each class began its drill work with the kind of material assigned. Both groups of classes devoted approximately the same time to this formal drill, ten minutes daily, such time being taken from the regular sixty minute daily period. The drill was always given at the beginning of the period. However, the pupils were allowed to use their practice sheets outside of class, if they wished, the slower ones in fact were even urged to do so. This is a perfectly legitimate use of practice material and is supposedly one of its merits. If such material has in it the elements of interest sufficient to stimulate pupils to a more diligent use, then just so much more valuable does such material become. In April the final Series B tests were given. The scores are given in the following tables:

*Charles W. Johnson, Williams Avenue School, Norwood; H. L. Crane, Oakley School, Cincinnati; D. S. Richards, Linwood School, Cincinnati.

TABLE I

Frequency Tables—Tests I and II (Series B)

OAKLEY SCHOOL — ATTEMPTS—TWO 5TH GRADES

Score	Courtis' Practice Pads								Thompson's Essentials							
	Addition		Subtrac.		Multip.		Division		Addition		Subtrac.		Multip.		Division	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
0	..	..	..	..	1	..	1	2	..	..	..	..	..	..	..	..
1	..	..	..	..	0	..	2	1	..	..	..	..	..	..	2	..
2	3	1	..	1	3	1	5	4	..	..	..	..	..	..	1	7
3	3	1	1	0	3	3	9	6	2	..	..	1	3	6	6	11
4	3	3	4	0	5	4	11	10	4	1	1	1	3	4	16	8
5	10	4	2	1	8	6	5	7	2	2	6	2	9	4	5	5
6	8	5	13	6	11	9	2	9	11	3	9	10	11	11	4	3
7	5	4	8	5	4	9	4	2	8	3	11	7	8	11	1	3
8	5	3	7	7	5	4	1	2	2	8	5	9	1	0	0	..
9	5	6	4	6	2	4	1	..	3	4	2	4	0	0	0	..
10	0	5	2	5	1	0	1	..	1	4	0	0	1	1	1	..
11	1	1	1	5	..	3	1	..	2	3	2	2	0	.	1	..
12	..	2	1	2	..	..	..	..	1	2	1	1	1	..	..	..
13	..	3	..	1	..	..	..	..	1	4	..	..	..	..	..	..
14	..	3	..	0	..	..	..	..	..	0	..	..	..	..	..	..
15	..	1	..	1	..	..	..	..	..	2	..	..	..	..	..	..
16	..	0	..	1	..	..	..	..	..	1	..	..	..	..	..	..
17	..	.	..	1	..	..	..	..	..	..	..	..	..	..	..	..
18	..	..	..	1	..	..	..	..	..	..	..	..	..	..	..	..
19	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
20	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	43	43	43	43	43	43	43	43	37	37	37	37	37	37	37	37
Median	6.4	9.2	7.3	9.3	6.2	6.9	4.5	4.9	7.0	9.5	7.2	7.7	6.3	6.5	4.6	4.1
Gain		2.8		2.0		.7		.4		2.5		.5		.2		—.5

OAKLEY SCHOOL — ACCURACY—TWO 5TH GRADES

Score	Courtis' Practice Pads								Thompson's Essentials							
	Addition		Subtrac.		Multip.		Division		Addition		Subtrac.		Multip.		Division	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
100	3	2	5	5	3	6	10	10	3	3	10	6	7	4	10	6
90	0	0	1	1	0	0	0	0	0	2	0	0	0	0	1	0
80	4	4	11	8	8	4	5	5	7	6	11	5	14	4	3	2
70	2	10	5	9	9	4	7	5	2	3	3	6	2	5	6	2
60	5	6	6	4	5	3	7	3	4	5	4	4	4	8	3	6
50	5	9	6	7	5	7	3	6	9	5	4	8	5	6	2	5
0—49	24	12	9	9	13	19	11	14	12	13	5	8	5	10	12	16
	43	43	43	43	43	43	43	43	37	37	37	37	37	37	37	37
Median	46	62	72	72	68	54	71	67	58	62	83	68	82	64	73	56
Gain		16		0		—14		—4		4		—15		—18		—17

TABLE II

Frequency Tables—Tests I and II (Series B)

LINWOOD-WILLIAMS AVENUE SCHOOLS — ATTEMPTS—TWO 5TH GRADES

Score	Courtis' Practice Pads								Thompson's Essentials							
	Addition		Subtrac.		Multip.		Division		Addition		Subtrac.		Multip.		Division	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
0	..	..	..	..	..	..	1	..	..	..	..	..	..	..	..	..
1	..	..	..	..	..	..	1	..	..	..	..	..	..	..	..	..
2	1	..	..	..	3	1	6	1	1	..	..	..	..	..	1	1
3	3	1	1	..	6	1	9	1	2	1	..	..	..	..	3	2
4	6	0	6	..	9	0	11	6	2	1	1	..	3	3	3	3
5	9	0	5	..	10	4	3	5	7	2	2	..	5	3	5	4
6	4	3	5	1	3	7	3	9	6	7	6	4	5	5	1	2
7	4	6	9	7	3	4	..	2	2	4	2	2	6	4	5	2
8	6	2	7	5	..	6	..	1	3	3	5	7	2	1	2	5
9	0	3	0	1	..	5	..	1	4	3	3	4	1	7	2	0
10	0	2	0	8	..	2	..	0	1	3	2	3	5	2	3	4
11	1	1	0	5	..	1	..	1	..	2	1	1	0	1	1	0
12	..	5	1	4	..	..	..	2	..	0	2	2	0	1	0	0
13	..	1	..	..	..	..	..	2	..	2	3	0	0	0	1	3
14	..	4	..	..	..	..	..	..	..	..	0	1	0	0	0	0
15	..	0	..	..	..	..	..	..	..	..	0	3	00	0	0	0
16	..	2	..	..	..	..	..	..	..	..	0	0	0	0	0	0
17	..	1	..	..	..	..	..	..	..	..	1	1	0	0	1	1
18	..	..	..	..	..	..	..	..	..	..	..	..	1	1	..	0
19	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1
20	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	34	31	34	31	34	31	34	31	28	28	28	28	28	28	28	28
Median	5.8	10.5	7.0	10.3	4.9	7.8	4.0	6.3	6.3	7.8	8.6	9.3	7.2	7.8	7.2	8.0
Gain		4.7		3.3		2.9		2.3		1.5		.7		.6		.8

LINWOOD-WILLIAMS AVENUE SCHOOLS — ACCURACY—TWO 5TH GRADES

Score	Courtis' Practice Pads								Thompson's Essentials							
	Addition		Subtrac.		Multip.		Division		Addition		Subtrac.		Multip.		Division	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
100	4	5	5	4	1	4	5	7	2	3	7	11	4	10	7	9
90	0	2	0	5	0	1	0	1	0	1	2	3	2	1	4	2
80	6	8	2	9	1	4	3	11	3	1	10	8	10	10	7	5
70	3	5	4	5	5	6	2	1	2	6	4	4	5	4	2	3
60	2	2	4	5	5	7	3	1	3	4	1	1	3	3	1	1
50	4	6	4	2	3	5	4	6	5	4	3	1	3	0	1	2
0—49	15	3	15	1	19	4	17	4	13	9	1	0	1	0	6	6
	34	31	34	31	34	31	34	31	28	28	28	28	28	28	28	28
Median	55	80	55	83	45	70	50	84	52	63	85	90	82	87	86	84
Gain		25		28		25		34		11		5		5		—2

TABLE III

Frequency Tables—Tests I and II (Series B)

WILLIAMS AVE.-LINWOOD SCHOOLS — ATTEMPTS—TWO 6TH GRADES

Score	Courtis' Practice Pads								Thompson's Essentials							
	Addition		Subtrac.		Multip.		Division		Addition		Subtrac.		Multip.		Division	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
0	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
1	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1	..
2	..	..	..	..	..	..	..	..	..	..	..	..	..	1	1	1
3	..	..	..	..	..	..	..	..	..	..	..	..	3	0	8	0
4	1		..	..	..	..	1	..	4	..	1	..	3	1	6	5
5	8	..	1	..	1	..	4	..	5	1	7	..	5	0	6	3
6	3	..	1	..	4	..	1	..	10	3	8	4	12	10	6	8
7	5	..	4	..	5	..	5	1	7	5	9	6	9	7	1	2
8	4	4	6	..	4	3	6	3	8	7	7	9	3	6	3	3
9	5	6	10	..	5	5	1	2	2	5	4	4	2	0	4	6
10	4	7	1	4	4	2	10	4	2	5	1	7	1	7	1	3
11	1	3	0	4	5	7	1	5	0	5	1	3	0	3	1	3
12	..	5	6	4	2	2	0	4	1	4	1	2	1	0	1	1
13	..	2	1	5	0	4	1	4	..	2	..	0	..	2	..	2
14	..	2	1	5	1	4	1	1	..	0	..	0	..	0	..	0
15	..	1	..	4	..	2	..	2	..	1	..	0	..	2	..	1
16	..	1	..	1	..	0	..	3	..	0	..	4	..	..	..	0
17	..	..	..	3	..	0	..	0	..	0	..	..	..	..	..	0
18	..	..	..	1	..	1	..	2	..	1	..	..	..	..	..	1
19	..	..	..	..	..	1	..	..	..	..	..	..	..	..	..	..
20	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	31	31	31	31	31	31	31	31	39	39	39	39	39	39	39	39
Median	7.8	10.8	9.4	13.8	9.4	11.8	8.8	12.3	7.1	9.8	7.4	9.3	6.8	8.1	5.7	8.3
Gain		3.0		4.4		2.4		3.5		2.7		1.9		1.3		2.6

WILLIAMS AVE.-LINWOOD SCHOOLS — ACCURACY—TWO 6TH GRADES

Score	Courtis' Practice Pads								Thompson's Essentials							
	Addition		Subtrac.		Multip.		Division		Addition		Subtrac.		Multip.		Division	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
100	2	3	3	12	6	5	14	11	2	10	3	4	4	7	10	10
90	0	8	3	7	4	4	2	10	0	1	1	4	0	3	0	0
80	7	4	11	6	6	8	8	4	3	13	13	6	10	9	5	11
70	4	8	5	1	5	8	4	2	4	5	5	8	0	4	4	5
60	6	1	2	3	6	4	2	0	10	3	5	7	10	5	5	3
50	6	4	5	1	2	1	1	3	7	2	6	4	5	8	2	3
0—49	6	3	2	1	2	1	0	1	13	5	6	6	10	3	13	7
	31	31	31	31	31	31	31	31	39	39	39	39	39	39	39	39
Median	67	80	82	96	82	83	95	96	60	84	76	74	65	80	70	82
Gain		13		14		1		1		24		—2		15		12

TABLE IV

ADDITION — SUBTRACTION

	Courtis			Thompson			Courtis			Thompson		
	Att.	Acc.	Eff.	Att.	Acc.	Eff.	Att.	Acc.	Eff.	Att.	Acc.	Eff.
Oakley 5th												
Test I	6.4	46	2	7.0	58	3	7.3	72	2	7.2	83	3
Test II	9.2	62	5	9.5	62	8	9.3	72	7	7.7	68	3
Change	2.8	16	3	2.5	4	5	2.0	0	5	.5	—15	0
Lin.–Wms. 5th												
Test I	5.8	55	6	6.3	52	4	7.0	55	0	8.6	85	11
Test II	10.5	80	13	7.8	63	7	10.3	83	10	9.3	90	25
Change	4.7	25	7	1.5	11	3	3.3	28	10	.7	5	14
Wms.–Lin. 6th												
Test I	7.8	67	3	7.1	60	3	9.4	82	0	7.4	76	0
Test II	10.8	80	10	9.8	84	15	13.8	96	39	9.3	74	3
Change	3.0	13	7	2.7	24	12	4.4	14	39	1.9	—2	3

MULTIPLICATION — DIVISION

	Courtis			Thompson			Courtis			Thompson		
	Att.	Acc.	Eff.	Att.	Acc.	Eff.	Att.	Acc.	Eff.	Att.	Acc.	Eff.
Oakley 5th.												
Test I	6.2	68	5	6.3	82	5	4.5	71	2	4.6	73	3
Test II	6.9	54	0	6.5	64	3	4.9	67	5	4.1	56	3
Change	.7	—14	—5	.2	—18	—2	.4	—4	3	—.5	—17	0
Lin.–Wms. 5th.												
Test I	4.9	45	0	7.2	82	7	4.0	50	0	7.2	86	11
Test II	7.8	70	6	7.8	87	14	6.3	84	13	8.0	84	25
Change	2.9	25	6	.6	5	7	2.3	34	13	.8	—2	14
Wms.–Lin. 6th.												
Test I	9.4	82	10	6.8	65	0	8.8	95	32	5.7	70	15
Test II	11.8	83	13	8.1	80	3	12.3	96	36	8.3	82	20
Change	2.4	1	3	1.3	15	3	3.5	1	4	2.6	12	5

For purposes of comparison, general (standard) median scores are inserted. Such medians have been determined by Courtis and are based upon the May and June, 1915–16 scores of many thousand pupils. Bulletin 4, p. 48, Courtis Standard Research Tests, Detroit.

Grade	Addition		Subtraction		Multiplication		Division	
	Sp.	Acc.	Sp.	Acc.	Sp.	Acc.	Sp.	Acc.
5—8.6	70		9.0	83	7.5	75	6.1	77
6—9.8	73		10.3	85	9.1	78	8.2	87

TABLE V

Frequency Tables—Tests I and II (Series B)

TOTAL—ALL PUPILS ATTEMPTS—ALL PUPILS

Score	Courtis' Practice Tests								Thompson's Essentials							
	Addition		Subtrac.		Multip.		Division		Addition		Subtrac.		Multip.		Division	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
0	..	..	..	..	1	..	2	2	..	..	..	..	..	..	..	..
1	..	..	..	..	0	..	3	1	..	..	..	..	..	..	3	..
2	4	1	..	1	6	2	11	5	1	..	..	..	..	1	3	9
3	6	2	2	0	9	4	18	7	4	1	..	1	6	6	17	13
4	10	3	10	0	14	4	23	16	10	2	3	1	9	8	25	16
5	27	4	8	1	19	10	12	12	14	5	15	2	19	7	16	12
6	15	8	19	7	18	16	6	18	27	13	23	18	28	26	11	13
7	14	10	21	12	12	13	9	5	17	12	22	15	23	22	7	7
8	15	9	20	12	9	13	7	6	13	18	17	25	6	7	5	8
9	10	15	14	7	7	14	2	3	9	12	9	12	3	7	6	6
10	4	14	3	17	5	4	11	4	4	12	3	10	7	10	5	7
11	3	5	1	14	5	11	2	6	2	10	4	6	0	4	3	3
12	..	12	8	10	2	2	0	6	2	6	4	5	2	1	1	1
13	..	6	1	6	0	4	1	6	1	8	3	0	0	2	1	5
14	..	9	1	5	1	4	1	1	..	0	0	1	0	0	0	0
15	..	2	..	5	..	2	..	2	..	3	0	3	0	2	0	1
16	..	3	..	2	..	0	..	3	..	1	0	4	0	0	0	0
17	..	2	..	4	..	0	..	0	..	0	1	1	0	0	1	1
18	..	..	..	2	..	1	..	2	..	1	..	..	1	1	..	1
19	..	..	..	..	..	1	..	..	..	..	..	..	..	..	..	1
20	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	108	105	108	105	108	105	108	105	104	104	104	104	104	104	104	104
Median	6.5	10.1	7.7	10.8	6.3	8.3	4.9	6.6	6.9	9.1	7.5	8.6	6.6	7.2	5.3	6.2
Gain		3.6		3.1		2.0		1.7		2.2		1.1		.6		.9

TOTAL—ALL PUPILS ACCURACY—ALL PUPILS

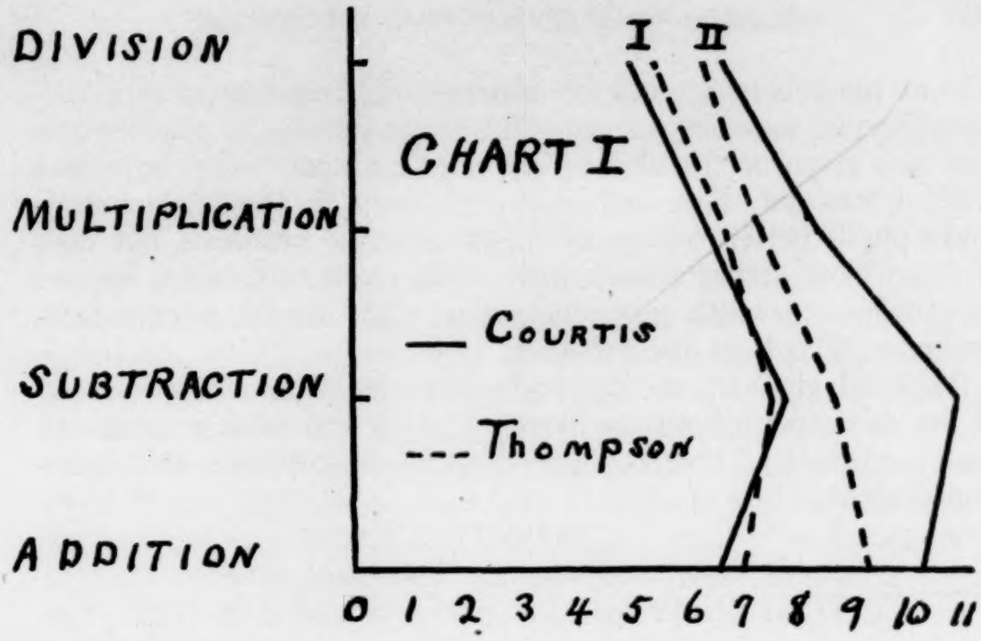
Score	Courtis' Practice Tests								Thompson's Essentials							
	Addition		Subtrac.		Multip.		Division		Addition		Subtrac.		Multip.		Division	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
100	9	10	13	21	10	15	29	28	7	16	20	21	15	21	27	25
90	0	10	4	13	4	5	2	11	0	4	3	7	2	4	5	2
80	17	16	24	23	15	16	16	20	13	20	34	19	34	23	15	18
70	9	23	14	15	19	18	13	8	8	14	12	18	7	13	12	10
60	13	9	12	12	16	14	12	4	17	12	10	12	17	16	9	10
50	15	19	15	10	10	13	8	15	21	11	13	13	13	14	5	10
0—49	45	18	26	11	34	24	28	19	38	27	12	14	16	13	31	29
	108	105	108	105	108	105	108	105	104	104	104	104	104	104	104	104
Median	56	73	71	82	66	71	75	84	57	71	81	77	79	77	76	73
Gain		17		11		5		9		14		—4		—2		—3

Table I gives the results in the fundamentals, preliminary (I) and final (II), with Series B Standard tests in two fifth grade classes of the Oakley school. It should be read: forty-three pupils of the class who practiced with the Courtis exercises made an initial median score in addition attempts of 6.4 examples, and a final median score in addition attempts of 9.2 examples, making a gain of 2.8 examples attempted. The parallel class of thirty-seven pupils who practiced with the Thompson exercises made a gain in addition attempts of 2.5 examples. The results in accuracy are read in the same manner. Tables II and III give similar data for the other paired classes. Table IV assembles the results by classes, with the score for efficiency added. Efficiency means the per cent. of pupils attempting or exceeding a certain standard number of examplesand working the same with one hundred per cent. accuracy. Some classes show a loss, as registered in the "change," in either number of attempts, or in accuracy, or efficiency.

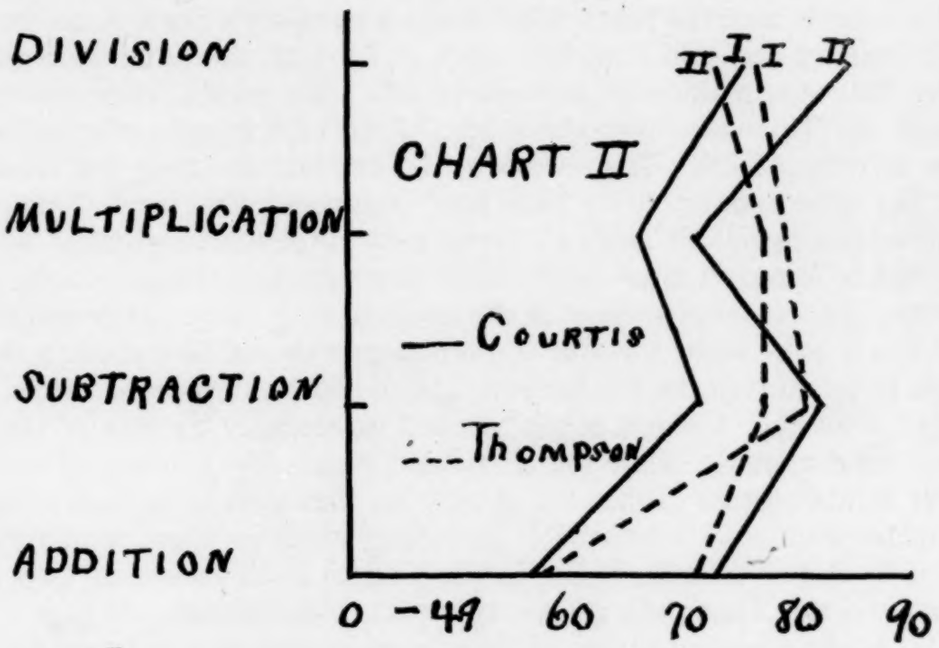
For purposes of comparison, the one hundred five pupils of the one set of exercises and the one hundred four pupils of the other set are grouped, regardless of school and grade, in Table V. These results are graphed in Charts I and II. It will be noticed that, in attempts, the pupils having the Courtis exercises made the greater increase in speed, the groups starting with about the same degree of rapidity. For accuracy of work, the Courtis group made a consistent gain in each of the fundamentals, and this based upon a greater increase in number of examples attempted. The Thompson group worked with greater accuracy at the initial test in each of the fundamentals, but, while making a consistent gain for addition, this group was less accurate at the final test in the three other processes. This shows in Chart II by the dotted lines crossing between addition and subtraction.

The Comparative Standing of the Two Groups in Initial And Final Tests in Arithmetical Reasoning

Parallel to and in conjunction with the above series of tests, data were obtained regarding any change in the ability of pupils to work so-called arithmetical reasoning problems. The Starch Scale A for arithmetical reasoning was given to all classes preliminary to and at the close of the practice drills. This scale, with tentative standards, may be found in the April, 1916, number of the *Journal of Educational Psychology*. Another set of ten problems was given



MEDIAN NUMBER OF EXAMPLES ATTEMPTED BY ALL PUPILS IN THE PRELIMINARY I AND FINAL II TESTS



MEDIAN PER CENT ACCURACY BY ALL PUPILS IN THE PRELIMINARY I AND FINAL II TESTS

also at the beginning, and for purposes of comparison, an approximately equal set weighted according to the number of pupils working, was given at the close. The question upon which light was wanted was not, does increased proficiency in the fundamentals make pupils better "reasoners" in arithmetical problems, but does it mean more accurate work when these fundamentals are applied to problems in which reasoning enters. The results in such tests are more difficult of interpretation.

Table VI gives the median preliminary and final scores for each of the classes with both the Starch Scale A and what we shall call the "combination" test, the second set of ten problems each mentioned above.

TABLE VI.

	Starch Arithmetical Scale A						Combination Test					
	C. 1	T. 1	C. 2	T. 2	C. 3	T. 3	C. 1	T. 1	C. 2	T. 2	C. 3	T. 3
Test I	5.0	7.7	4.0	7.4	11.3	7.6	8.9	9.1	11.4	18.0	56.0	34.8
Test II	7.2	9.5	7.5	10.0	13.3	10.0	27.8	41.0	33.6	46.5	72.4	54.6
Change	2.2	1.8	3.5	2.6	2.0	2.4	18.9	31.9	22.2	26.5	16.4	19.8

This table should be read: one of the classes which practiced with the Courtis material made a 5.0 median score with Scale A on the preliminary test and a median score of 7.2 with the same scale on the final test, making an increase of 2.2. The parallel class which used the Thompson material made 7.7 and 9.5 respectively, with an increase of 1.8. The results in the combination tests are read in the same manner. We have here "compared class results" similar to those given in Table IV in the tests in the fundamentals. As would be expected after five months' classroom instruction in arithmetic, there is improvement in the scores in each class. How much of this improvement was due to normal growth and how much was due to practice in the fundamentals is, of course, uncertain. However, since the "Courtis pupils" gained in accuracy in each of the four fundamentals, while the Thompson pupils lost in three of the four fundamentals (Table V), it may be that gain in accuracy in fundamentals may account for the rather favorable showing of the Courtis classes with Scale A. In the Combination reasoning tests, however, the Thompson groups appeared to advantage.

So long as median initial abilities of groups to be compared are not approximately equal, just what constitutes greater improvement is difficult to determine. Is a change from 5.0 to 7.2 better than a

change from 7.7 to 9.5? The first class has "more room" for improvement. These results with reasoning problems are given for what they might be considered worth.

SUMMARY

One hundred five pupils practicing with one kind of practice material ten minutes daily for ninety school days, such time being taken out of the regular sixty minute arithmetic period, made a greater improvement in work in the fundamentals than did another group of one hundred four pupils under similar conditions with another kind of practice material. There was a noticeable improvement in the ability to work problems involving arithmetical reasoning in all classes regardless of practice material. The results, with regard to the fundamentals, substantiate the findings of a former experiment with nearly one thousand pupils.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except July and August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Board of Editors

WILLIAM CHANDLER BAGLEY, CARL E SEASHORE,
J. CARLETON BELL, Managing Editor, GUY MONTROSE WHIPPLE.

——COLLABORATORS——

Sydney Alrutz,	H. H. Goddard,	Will S. Monroe,	L. William Stern
J. R. Angell,	Joseph Jastrow	E. C. Moore,	G. M. Stratton,
J. W. Baird,	G. E. Johnson,	M. V. O'Shea,	G. D. Strayer,
Bird T. Baldwin,	E. A. Kirkpatrick,	W. B. Pillsbury,	E. L. Thorndike,
W. Van Dyke Bingham,	E. H. Lindley,	S. H. Rowe,	E. B. Tichener,
E. F. Buchner,	Anna J. McKeag,	Sante de Sanctis,	J. B. Watson,
W. H. Burnham,	D. P. MacMillan,	E. G. Sanford,	W. H. Winch,
Ed. Claparede,	Adolf Meyer,	W. S. Small,	H. K. Wolfe,
S. S. Colvin,	J. B. Miner,	E. D. Starbuck,	A. H. Yoder.
W. F. Dearborn,			

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated.

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 per year of ten monthly issues, if paid in advance. Single issues are 50 cents each. Back volumes are $3.50 in original wrappers and $4.75 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME IX MAY, 1918 NUMBER 5

EDITORIAL

War and Credulity

While it is the proverbial last straw that is supposed to break the camel's back, most breaks occur not through the successive increments of straws, but as the result of internal stresses and strains which were unrecognized until some sudden increase in the burden revealed the weakness. An instance of this is supplied in the phenomenon of "shell shock" of which we hear so much now-a-days. The name appears to be something of a misnomer, for the condition has none of the suddenness of onset connoted by the term "shock," nor is it caused by the bursting of any particular shells. It is rather the general result of the intense strain of the firing line upon nervous organizations that are temperamentally predisposed to be affected by horrible experiences. The massing of such experiences on the susceptible individual before the organism is able to react to them by habituation produces an overpowering, deadening, stupefying effect which may result in complete mental tetanus. The large majority of the men in the firing line, if they have specific work to do, react quickly to the situation and are able to throw off the depression.

In like manner the general strain and anxiety of war reveals the latent weaknesses of the civilian community. One of the most striking of these is the greatly enhanced credulity with which even intelligent people receive the wildest and most improbable statements. One needs only to recall the stories of infected court plaster, of poisoned food, of broken glass in candy and bread, which have had such wide credence in the past year. Dwellers in Atlantic seaboard cities know with what awed fascination preposterous tales of German air raids along the American coast, of wholesale transport sinkings, of hundreds of thousands of American wounded secretly brought back to this country and crowded into local hospitals, have been eagerly imbibed and passed on to others.

A particularly sinister form of this weakness of credulity seems to be on the increase among the intellectuals of England and France in the acceptance of evidences of spiritistic manifestations. With the increasing losses of friends and relatives and the general gloom occasioned by the war, there is a quickening of religious feeling, and a readiness to lend a willing ear to stories of communication with the spirit world. This credulity is especially difficult to combat because of the impossibility of proving the negative, because of the strong emotional bias in its favor, and because of its agreement with the tenor of religious dogmas implicitly believed from childhood. No one was greatly surprised when Sir Oliver Lodge professed to have received communications from the spirit of his son, Raymond, who had been killed in the war, for Sir Oliver's belief in spiritism has long been professed. But now comes Sir Arthur Conan Doyle, backed by his prestige as a popular writer of fiction, declaring his conviction, reached since the outbreak of the war, that spiritism is the new revelation of truth to humanity, "the rebirth of the original faith and practice of Christ and his immediate followers." Likewise, we have the testimony of W. J. Crawford, professor of mechanical engineering in Queen's College, Belfast, in a book in which he claims to have applied to spiritistic phenomena "the same principles and methods of investigation and test that he would use in his laboratory upon the manifestations of any accepted and understood form of force." Finally, there has recently been translated into English a book called *The Psychology of the Future*, by the French pseudo-psychologist Emile Boirac, who before his death last year had suffered the almost complete destruction of his family by the war. M. Boirac seems to have been a very credulous person, and professed his faith in all sorts of occult phenomena, as spiritism, thought transference, vital radiation, and other manifestations of psychic energy. He even contends that this way lies the path of scientific psychology. Such a wreckage of rigorous logical thinking under the emotional stress of the war would be merely pitiful, if it were not for the vicious impulse to credulity which this flood of occult literature from men who will be quoted as authoritative will give to the uncritical masses. From this point of view their influence is most pernicious.

As G. Stanley Hall truly says in that altogether delightful study *A Medium in the Bud*, "Psychic researchers today represent the last potent stand of about all the old superstitions of the past against which science has contended. The next generation will be hardly able to believe that prominent men in this wasted their energies in chasing such a will-o'-the-wisp as the veracity of messages or the reality of a post-mortem existence, which they no more prove than dreams of levitation prove that men can hover in the air at will." If education is to make for straight thinking and right living, specific measures should be taken by teachers even in the elementary school to make their pupils realize how absurd, illogical and superstitious such beliefs are. In the psychological courses to prospective teachers throughout the country a special effort should be made to hold such antiquated medievalism up to ridicule, and thus to work for immunity against this particular disease of the social consciousness.

J. C. B.

NOTES AND NEWS

A group of students of man have formed a "Galton Society" at New York City. The charter members are Madison Grant, George S. Huntington, Charles B. Davenport, C. H. Merriam, William K. Gregory, J. H. McGregor, Edward L. Thorndike, Edwin G. Conklin and H. F. Osborn. It is proposed to elect members from time to time up to twenty-five in number. Dr. Davenport was elected president and Dr. Gregory secretary.

"A Scale for Measuring Ability of Children in Geography in Grades IV–VIII" has been published by H. H. Hahn and E. E. Lackey, of the Wayne State Normal School, Wayne, Nebraska. Answers were obtained to 225 selected geography questions from 1696 pupils in 12 schools. The scale is arranged on the plan of the Ayres Spelling Scale, the questions being divided into twenty-five groups, from the easiest to the hardest, and the percentage made by each school grade on each group being indicated at the top. The authors have promised a more extended account of the derivation of the scale for an early number of the *Journal*.

Mr. Arthur S. Otis has recently issued the printed forms and manual of directions for his Absolute Point Scale, an account of which appears in the present and following issues of the *Journal*. The price will be $10.00 per hundred sets, including the various record sheets. The tests are issued in two forms, A and B, so that the testing can be duplicated at short intervals without any vitiating memory factor.

The Iowa Child Welfare Research Station has issued an individual record blank for recording the achievement of a child in the Terman Scale and the Yerkes Point Scales for six consecutive years. This device, facilitates a rapid survey of the child's intellectual development in successive years.

The Bureau of Educational Measurements and Standards of the Kansas State Normal School has issued an interesting pamphlet setting forth the values of standardized tests to the superintendent, the teacher, and the pupil, and giving a brief description of the more important tests, with prices for the blanks and a reference to the most complete discussion of the test.

The Amsterdam Paidological Association, under the presidency of Dr. G. A. M. Van Wayenburg, has established a new educational journal under the title of "*Kinderstudie.*" The journal is issued at irregular intervals, and the subscription price to foreigners is 4.50 florins (approximately $2.00) per volume. The journal will deal with all phases of child development, especially the development of intelligence.

The publishers of "The World Book," Chicago, are issuing an attractive and stimulating monthly 'service bulletin' for the benefit

to teachers. The bulletin contains questions on topics pertaining of school work, and is constructed with the view to arouse an attitude of inquiry toward timely items of general information.

On Friday night, April 26, three of the cottages at Dr. Maximilian P. E. Groszmann's institution for exceptional children, "Watchung Crest," Plainfield, New Jersey, were destroyed by fire. The main building was left uninjured. One of the cottages contained Dr. Groszmann's library and private clinic, and he had the misfortune to lose his entire collection of books, manuscripts, records, notes and correspondence.

The Berkshire Industrial Farm, at Canaan, N. Y., has established a department of clinical psychology under the supervision of Dr. Clinton P. McCord, of Albany. It is planned to have a complete physical, neurological and psychological examination of each boy in the institution, and this will be followed by regular monthly tests thereafter.

Professor Samuel P. Hayes, of Mt. Holyoke College, will appreciate any gifts of psychological literature to help replenish the losses sustained in the burning of the psychological library on December 20.

Among the lecturers from other institutions on the faculty of the summer sesssion of the School of Education, University of Chicago, are William Anton Schmidt, professor of education in the University of Oklahoma, Leonard V. Koos, associate professor of education in the University of Washington, and Edward Herbert Cameron, assistant professor of education in Yale University.

Dr. Thorstein B. Veblen, of the University of Missouri, author of *Theory of the Leisure Class* and *The Instinct of Workmanship*, has been appointed professor of economic institutions in Cornell University.

Miss Katharine B. Graves has been appointed instructor in educational psychology in the Woman's College of Delaware, Neward, Delaware.

Dr. George V. N. Dearborn has been commissioned a firstl ieutenant in the Medical Reserve Corps.

Dr. R. B. Teachout, instructor in psychology in the University of Oregon, has entered the psychological service of the national army.

Lieutenant M. R. Trabue has been promoted to captain and assigned to the officers' training school for psychological examiners at Camp Greenleaf, Fort Oglethorpe, Ga.